UNIWIN VERSION 10.2.0

CLASSIFICATION PAR LA METHODE DES K-MEDOIDES

Révision : 25/03/2025

Définition

La classification par la méthode des K-médoïdes est une approche de classification apparentée à la méthode des K-moyennes pour partitionner un ensemble de données en k classes. Dans la classification par K-médoïdes, chaque classe est représentée par l'une des observations de la classe. Ces observations sont appelées médoïdes.

Le terme médoïde fait référence à une observation au sein d'une classe pour laquelle la dissemblance moyenne entre elle et toutes les autres observations de la classe est minimale. Elle correspond au point le plus central de la classe. Ces observations (une par classe) peuvent être considérées comme des exemples représentatifs des membres des classes. Rappelons que, dans la classification par K-moyennes, le centre d'une classe donnée est calculé comme la valeur moyenne de toutes les observations de cette classe.

La méthode des K-médoïdes est une alternative robuste à la méthode des K-moyennes. Cela signifie que l'algorithme est moins sensible au bruit et aux valeurs aberrantes car il utilise les médoïdes comme centres des classes au lieu des moyennes.

La méthode des K-médoïdes la plus courante est l'algorithme PAM (Partitioning Around Medoids) de Kaufman et Rousseeuw (1990).

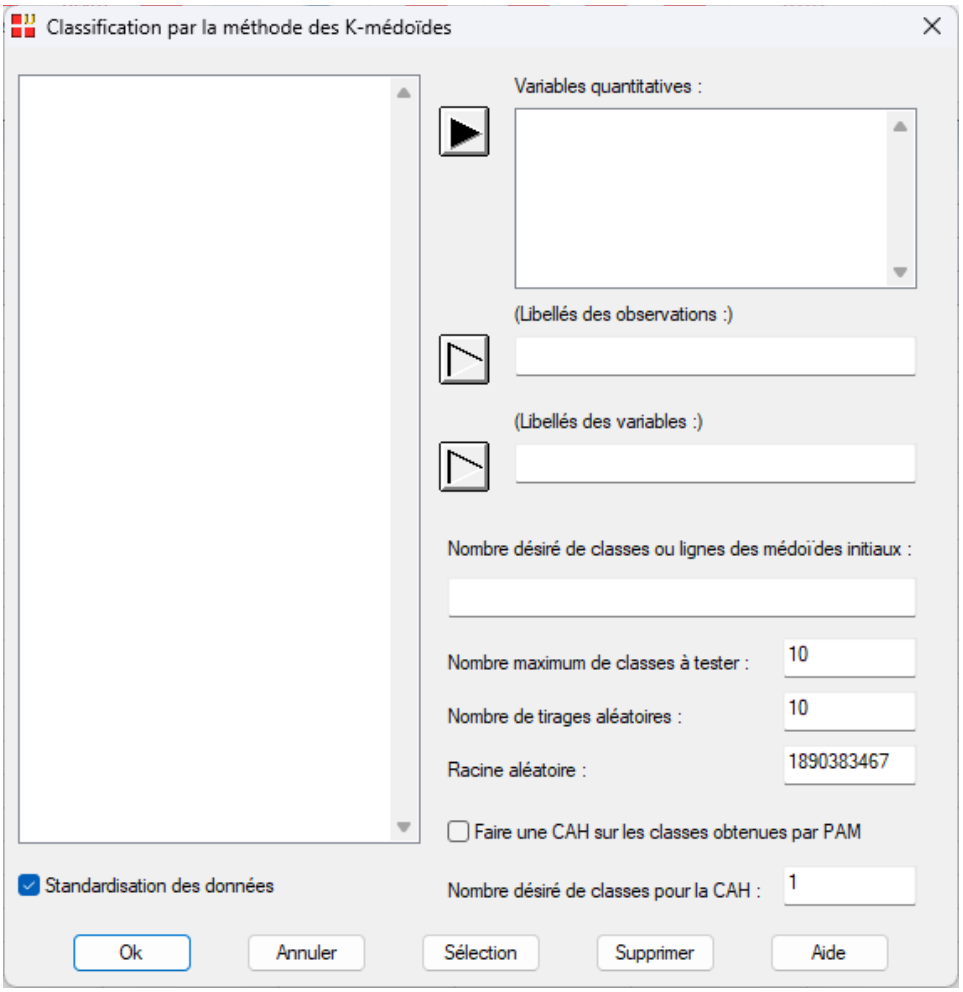
La procédure affiche un rapport indiquant notamment les médoïdes des classes formées, la classification des observations, des statistiques descriptives pour les classes formées et les contributions des variables aux classes. Si une classification mixte a été mise en œuvre, les résultats de la CAH sont également fournis.

Les graphiques des coefficients moyens et individuels de silhouette et des nuages de points des classes formées sont proposés. Si une classification mixte a été mise en œuvre, le diagramme des indices de la classification et l'affichage de l'arbre sont proposés.

Cette procédure est basée sur les packages R 'stats' et 'cluster'.

Entrée des données

Cliquons sur l'icône KM dans le ruban Décrire et choisissons K-médoïdes pour afficher la boîte de dialogue montrée ci-dessous :



Cette boîte de dialogue permet de choisir les variables quantitatives à utiliser pour la classification, la variable contenant les libellés des observations et la variable contenant les libellés des variables quantitatives utilisées pour la classification.

Par défaut les données sont standardisées par soustraction de la moyenne et division par l'écart absolu moyen (MAD).

Le nombre désiré de classes peut être précisé de deux façons, soit en entrant ce nombre, soit en entrant les numéros des lignes du fichier des données définissant les médoïdes initiaux de ces classes.

Le champ 'Nombre maximum de classes à tester' permet de calculer diverses statistiques pouvant aider à valider le nombre adéquat de classes à former.

Le nombre maximum d'itérations de l'algorithme peut être précisé.

Le nombre de tirages aléatoires des médoïdes de départ et la racine aléatoire associée peuvent être indiqués dans le cas où le nombre de classes à former est précisé. Ces deux options ne sont pas utilisées si les médoïdes initiaux des classes à former sont entrés.

Enfin, il est possible de compléter l'analyse par une classification ascendante hiérarchique (CAH) sur les classes obtenues par les K-médoïdes.

Données manquantes

Les données manquantes ne sont pas autorisées par cette procédure.

Exemple 1 : Fichier PAM

Pour ce premier exemple, nous utiliserons le fichier PAM.

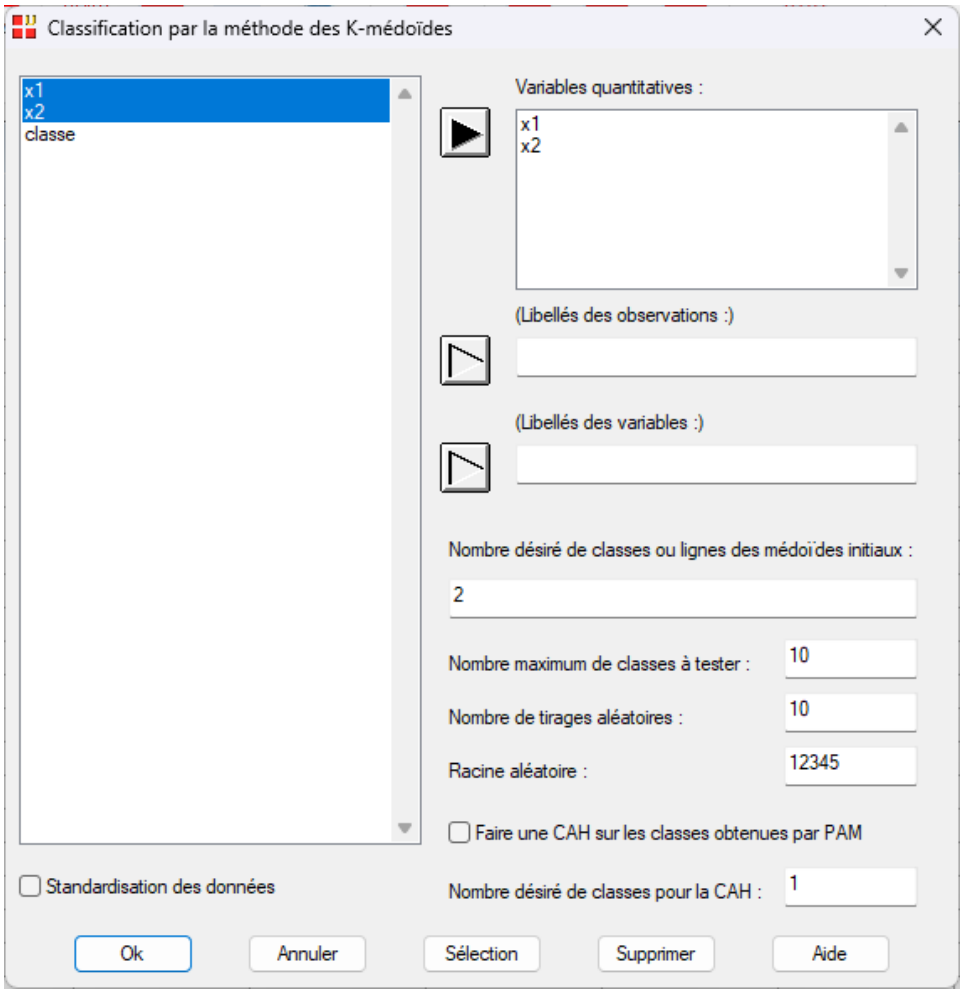
Il contient 25 observations (10 pour la classe 1 et 15 pour la classe 2) et deux variables x1 et x2 dont les données sont issues de tirages aléatoire de lois normales.

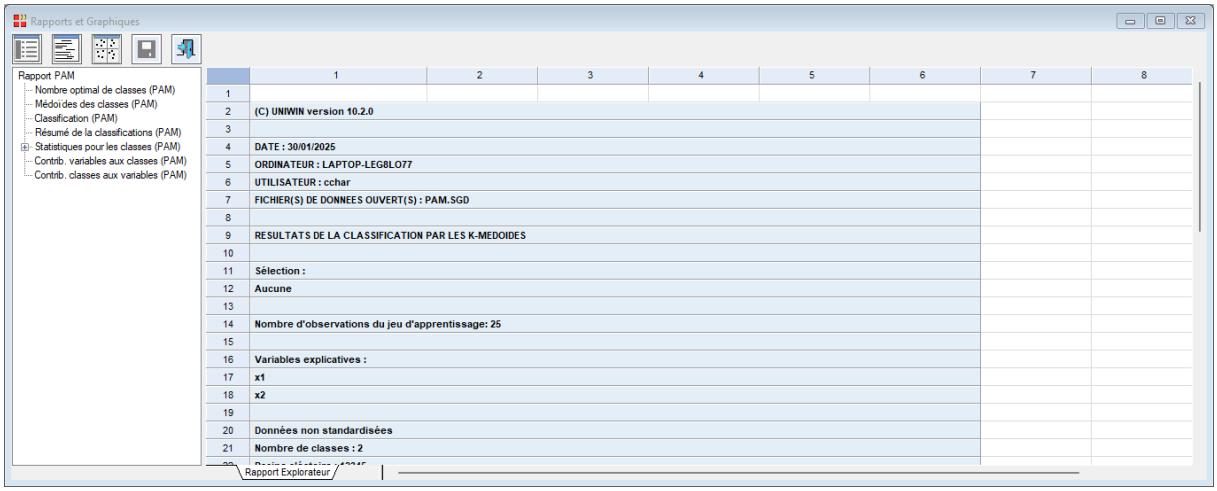
Cliquons sur l'icône KM dans le ruban Décrire et choisissons K-médoïdes et renseignons la boîte de dialogue montrée ci-après.

Nous demandons 2 classes et de faire une classification sur les données non standardisées.

Cliquons sur Ok. UNIWIN débute le calcul de la classification.

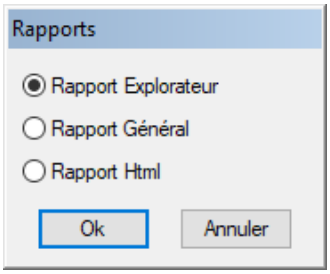
Après quelques instants, l'écran suivant s'affiche :



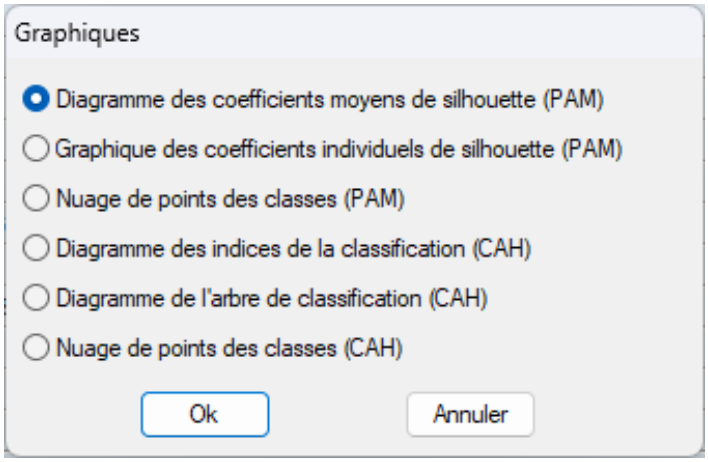


La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

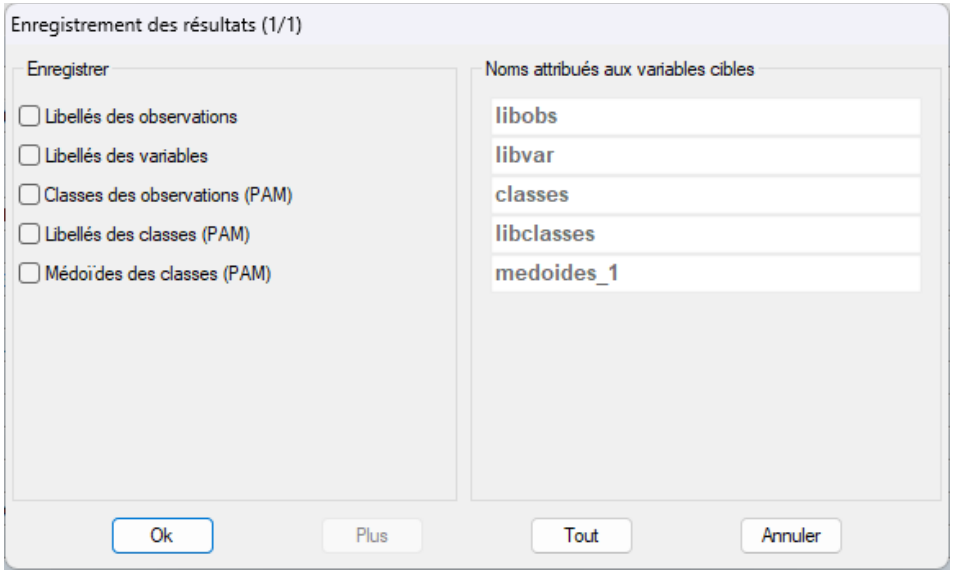
L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



et l'icône 'Graphiques'  affiche la boîte de dialogue des options pour les graphiques.



L'icône 'Enregistrer'  permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

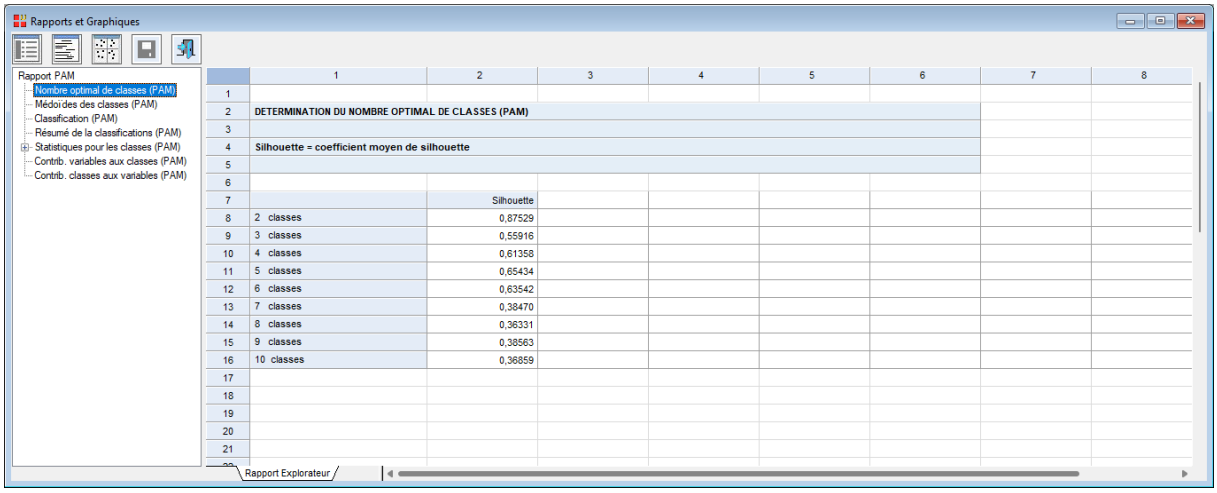


L'option Rapports

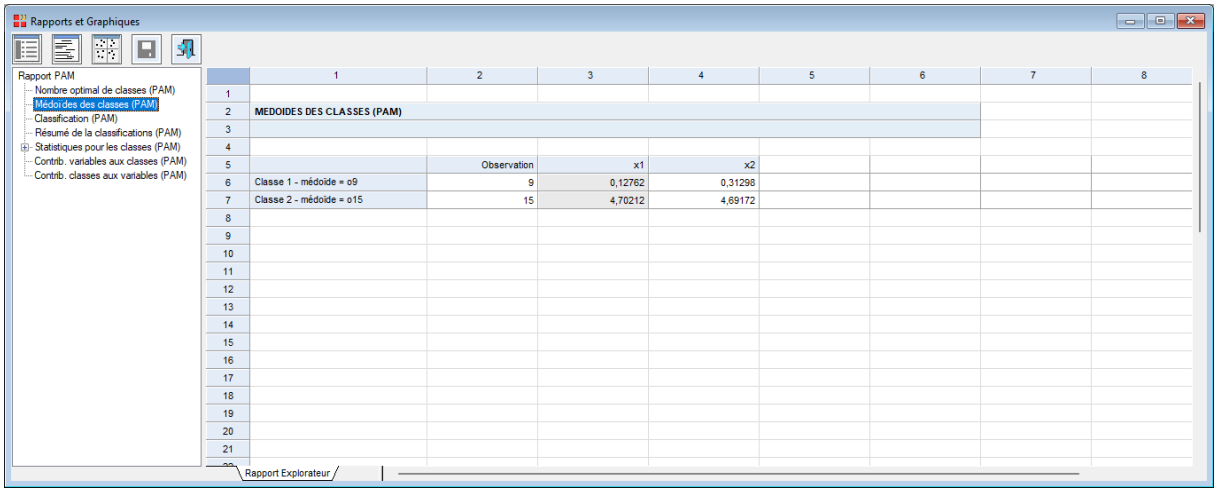
Cette option permet d'obtenir le rapport à l'écran sous la forme d'un tableur ou au format HTML.

Le premier tableau affiche pour les nombres de classes variant de 2 au nombre maximum de classes indiqué (10), les coefficients moyens de silhouette.

Ces coefficients sont également représentés de façon graphique. Ils aident à déterminer le nombre adéquat de classes à former.

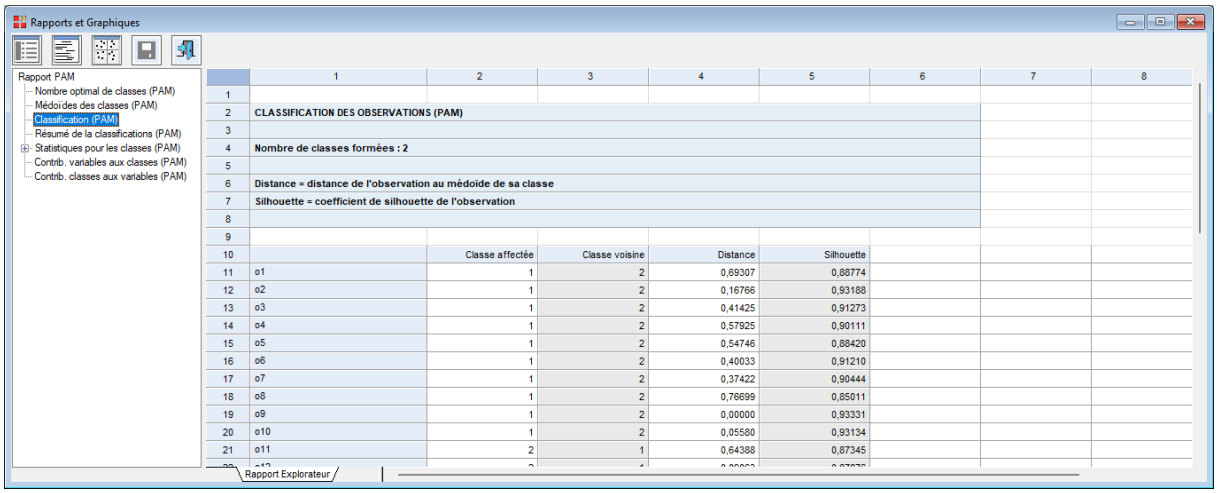


Le deuxième tableau affiche les médoïdes des classes formées.



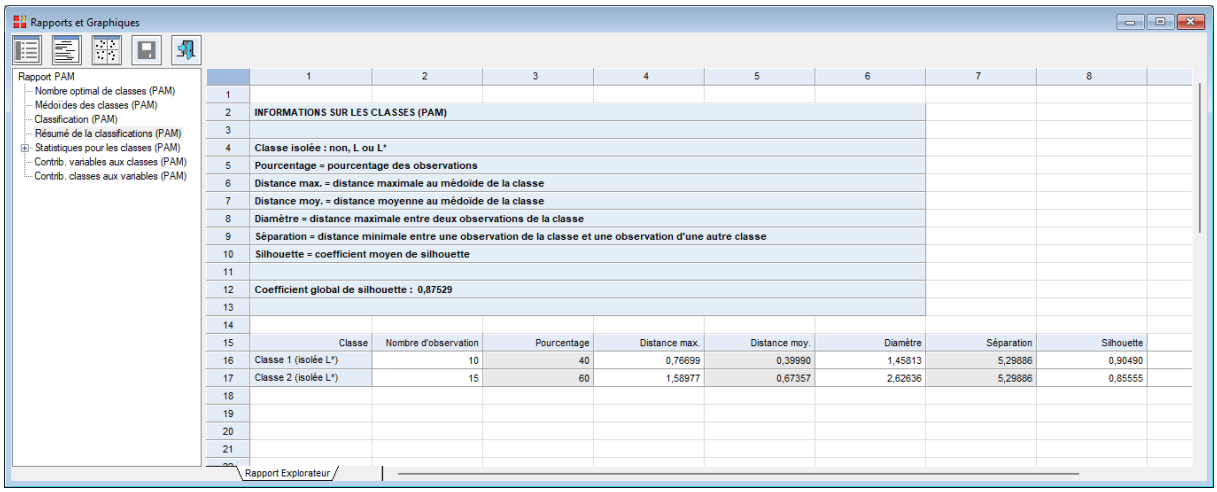
La classe 1 est représentée par l'observation 9 et la classe 2 par l'observation 15.

Le troisième tableau affiche pour chaque observation sa classe d'affectation, sa classe voisine, sa distance au médoïde et son coefficient de silhouette.



	Classe affectée	Classe voisine	Distance	Silhouette
o1	1	2	0,69307	0,88774
o2	1	2	0,16766	0,93188
o3	1	2	0,41425	0,91273
o4	1	2	0,57925	0,90111
o5	1	2	0,54746	0,88420
o6	1	2	0,40033	0,91210
o7	1	2	0,37422	0,90444
o8	1	2	0,76699	0,85011
o9	1	2	0,00000	0,93331
o10	1	2	0,05580	0,93134
o11	2	1	0,64388	0,87345

Le quatrième tableau affiche un résumé de la classification.

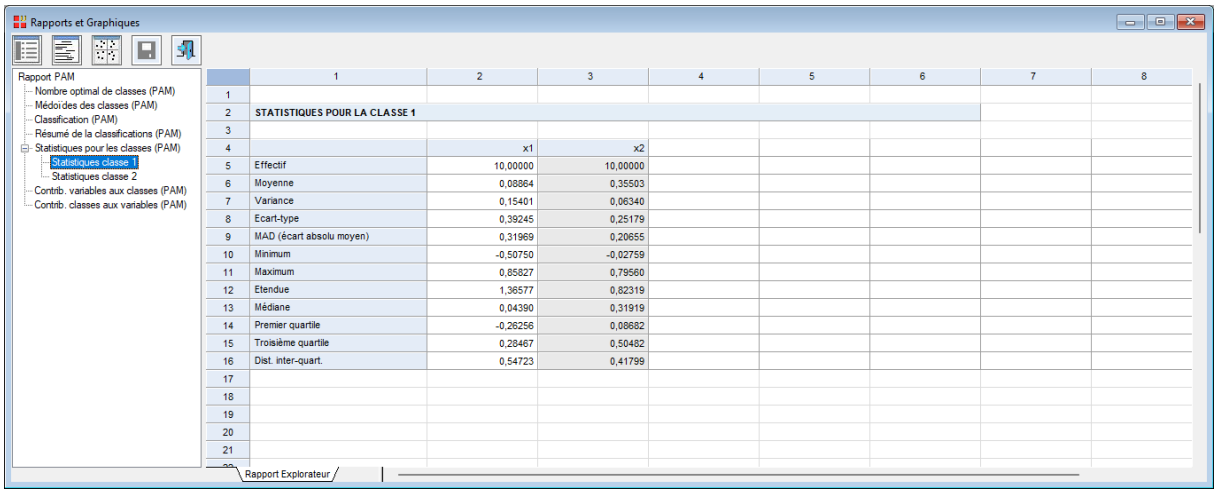


Classe	Nombre d'observation	Pourcentage	Distance max.	Distance moy.	Diamètre	Séparation	Silhouette
Classe 1 (isolée L*)	10	40	0,76699	0,39990	1,45813	5,29886	0,90490
Classe 2 (isolée L*)	15	60	1,58977	0,67357	2,62636	5,29886	0,85555

- Classe isolée (non isolée, L ou L*)
- Pourcentage des observations
- Distance maximale au médoïde de la classe
- Distance moyenne au médoïde de la classe
- Diamètre de la classe : distance maximale entre deux observations de la classe
- Séparation de la classe : distance minimale entre une observation de la classe et une observation d'une autre classe
- Coefficient moyen de silhouette

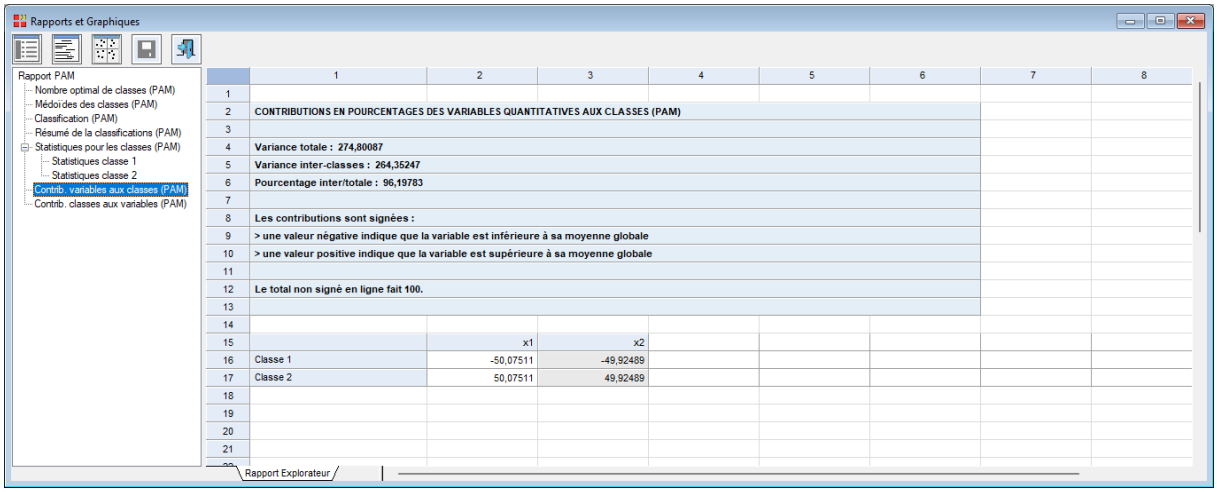
Une classe est une classe L* si son diamètre est plus petit que sa séparation. Une classe est une classe L si pour chaque observation i la dissimilarité maximale entre i et toute autre observation du cluster est plus petite que la dissimilarité minimale entre i et toute observation d'un autre cluster. Une classe L* est donc également une classe L.

Le cinquième tableau affiche pour chaque classe un ensemble de statistiques descriptives pour chacune des variables utilisées pour la classification.

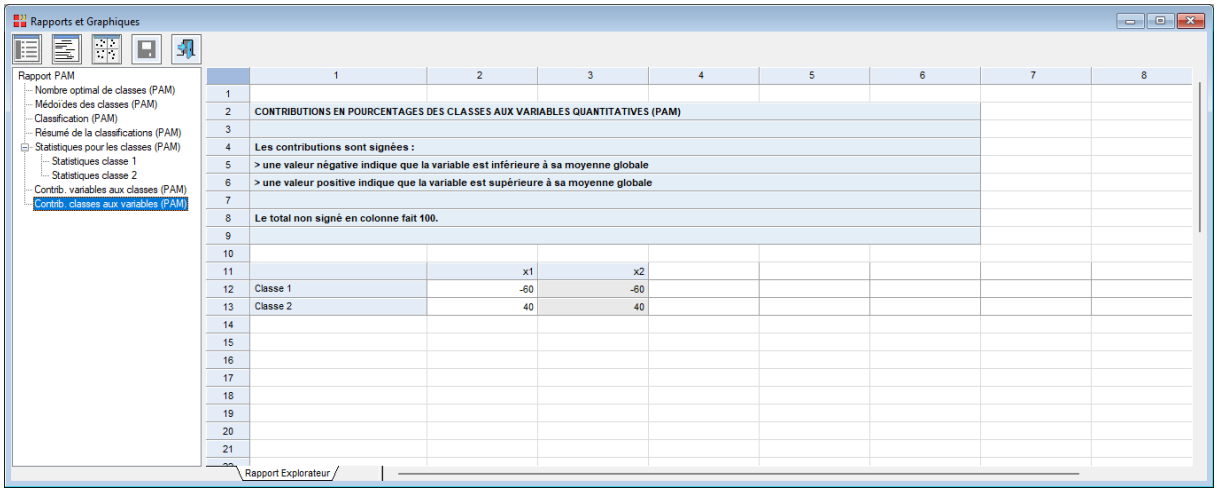


	1	2	3	4	5	6	7	8
1								
2	**STATISTIQUES POUR LA CLASSE 1**							
3								
4		x1	x2					
5	Effectif	10,00000	10,00000					
6	Moyenne	0,08864	0,35503					
7	Variance	0,15401	0,06340					
8	Ecart-type	0,39245	0,25179					
9	MAD (écart absolu moyen)	0,31969	0,20655					
10	Minimum	-0,50750	-0,02759					
11	Maximum	0,85827	0,79560					
12	Etendue	1,36577	0,82319					
13	Médiane	0,04390	0,31919					
14	Premier quartile	-0,26256	0,08682					
15	Troisième quartile	0,28467	0,50482					
16	Dist. inter-quart.	0,54723	0,41799					
17								
18								
19								
20								
21								

Les deux tableaux suivants affichent les contributions signées en pourcentages des variables aux classes et les contributions signées en pourcentages des classes aux variables.



	1	2	3	4	5	6	7	8
1								
2	**CONTRIBUTIONS EN POURCENTAGES DES VARIABLES QUANTITATIVES AUX CLASSES (PAM)**							
3								
4	**Variance totale : 274,80087**							
5	**Variance inter-classes : 264,35247**							
6	**Pourcentage inter/totale : 96,19783**							
7								
8	**Les contributions sont signées :**							
9	**> une valeur négative indique que la variable est inférieure à sa moyenne globale**							
10	**> une valeur positive indique que la variable est supérieure à sa moyenne globale**							
11								
12	**Le total non signé en ligne fait 100.**							
13								
14								
15		x1	x2					
16	Classe 1	-50,07511	-49,92489					
17	Classe 2	50,07511	49,92489					
18								
19								
20								
21								

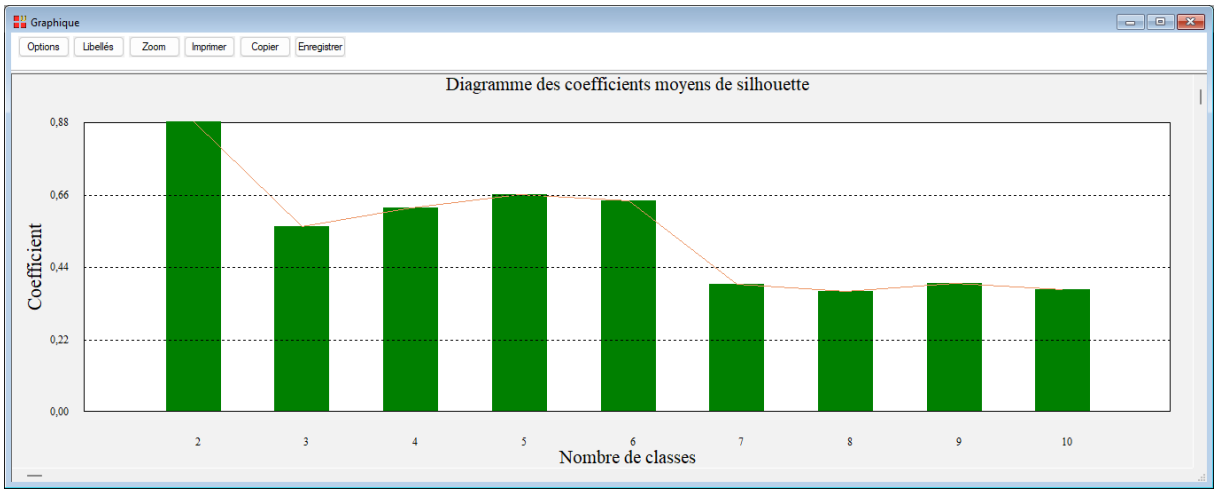


	1	2	3	4	5	6	7	8
1								
2	**CONTRIBUTIONS EN POURCENTAGES DES CLASSES AUX VARIABLES QUANTITATIVES (PAM)**							
3								
4	**Les contributions sont signées :**							
5	**> une valeur négative indique que la variable est inférieure à sa moyenne globale**							
6	**> une valeur positive indique que la variable est supérieure à sa moyenne globale**							
7								
8	**Le total non signé en colonne fait 100.**							
9								
10								
11		x1	x2					
12	Classe 1	-60	-60					
13	Classe 2	40	40					
14								
15								
16								
17								
18								
19								
20								
21								

L'option Graphiques

- Diagramme des coefficients moyens de silhouette

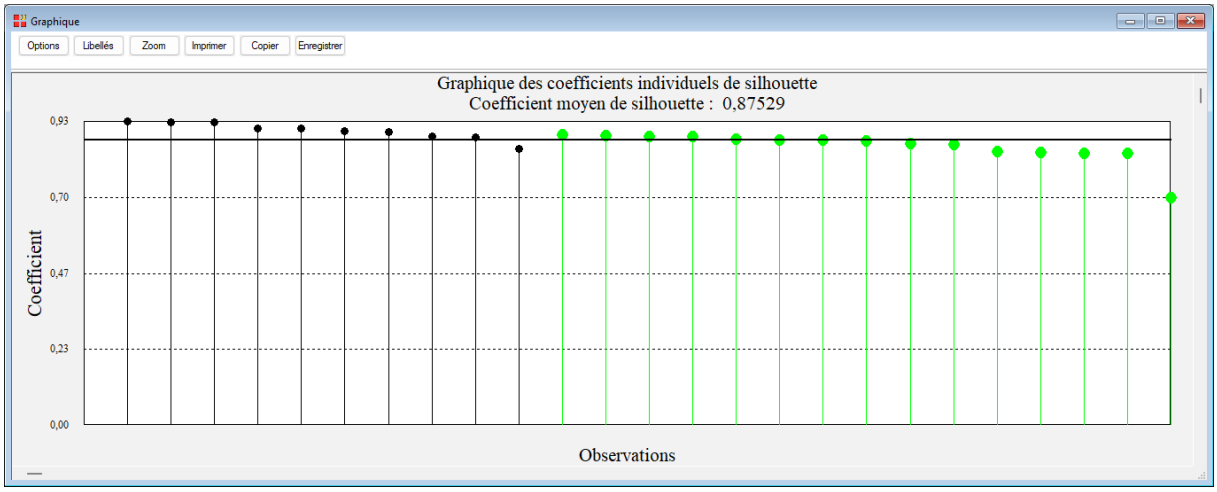
Le diagramme des coefficients moyens de silhouette permet de visualiser l'évolution de ces coefficients en fonction du nombre de classes formées.



- Graphique des coefficients individuels de silhouette

Le graphique des coefficients individuels de silhouette permet de visualiser ces coefficients pour chacune des observations. Pour chaque observation, le coefficient de silhouette est la différence entre la distance moyenne avec les observations de sa classe et la distance moyenne avec les observations des autres classes. Il varie entre -1 et +1. Si négatif, l'observation est en moyenne plus proche de la classe voisine que de la sienne et donc elle est donc mal classée. Si positif, l'observation est en moyenne plus proche de sa classe que de la classe voisine et donc elle est donc bien classée.

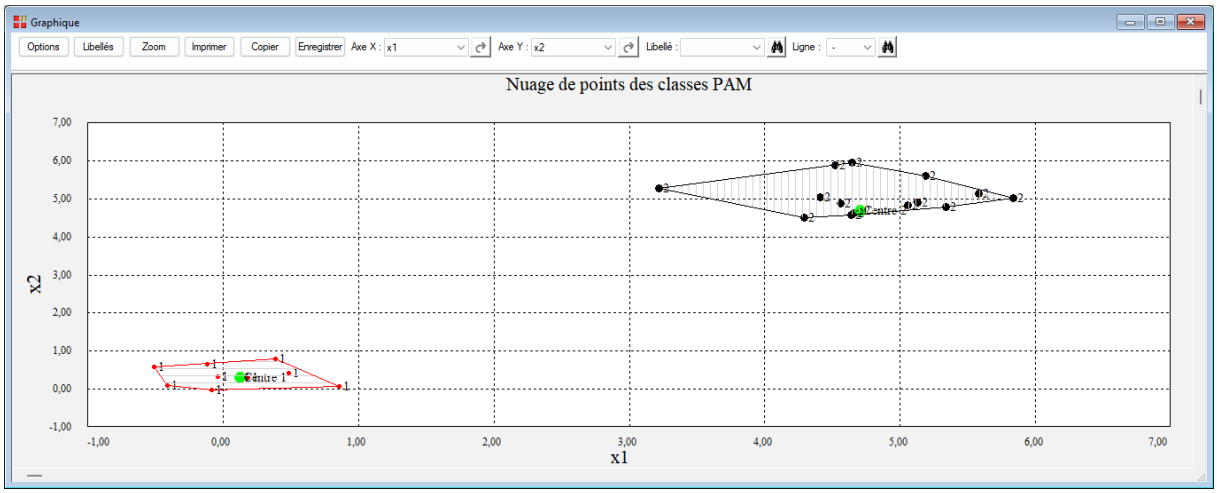
La ligne horizontale indique la position du coefficient moyen de silhouette. Si la plupart des observations d'une classe ont des coefficients inférieurs à ce coefficient moyen, cela indique probablement que la partition obtenue n'est pas bonne.

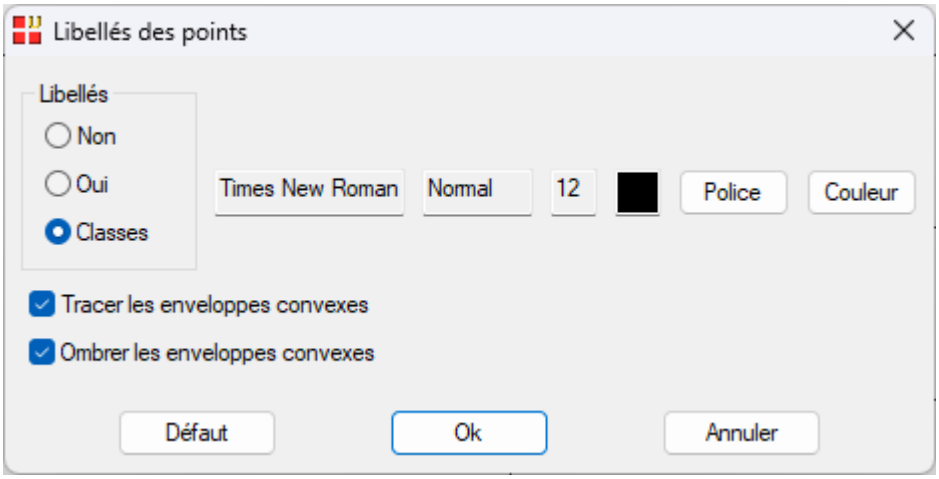


- Nuage de points des classes

Le graphique du nuage de points des classes permet de visualiser les classes formées par rapport à deux variables quantitatives sélectionnées.

Le bouton ‘Libellés’ permet de préciser les libellés affichés et si les enveloppes convexes des classes sont tracées ou non.





Exemple 2 : Fichier VEHICULE

Pour illustrer cet exemple, nous utiliserons le fichier VEHICULE.

Ce fichier contient 7 informations caractérisant 24 véhicules : *Modèle, Cylindrée, Puissance, Vitesse, Poids, Longueur, Largeur.*

Les libellés des variables quantitatives sont dans la variable *Mesures.*

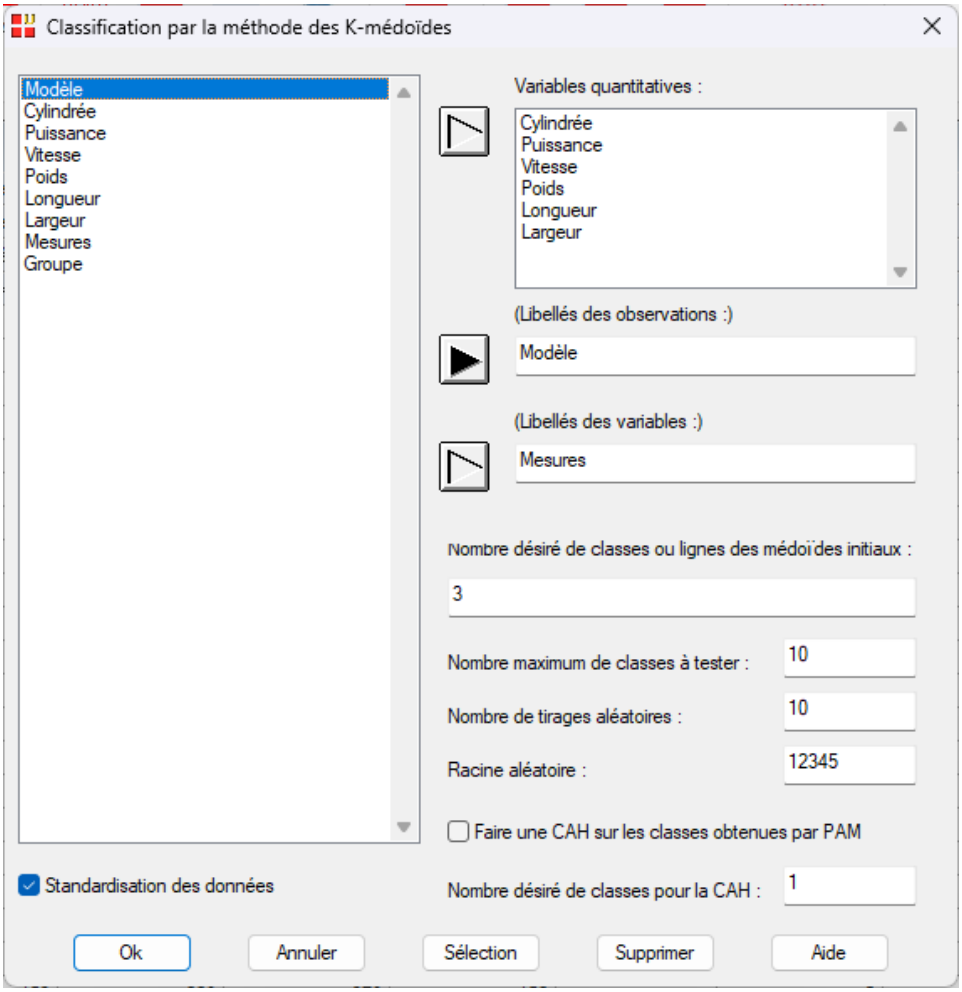
Les 24 modèles d'automobiles sont :

Honda Civic	R19	Fiat Tipo	405
R21	BX	BMW 530i	Rover 827i
R25	Opel Omega	405 Break	Ford Sierra
BMW 325ix	Audi 90 Quattro	Ford Scorpio	Espace
Nissan Vanette	VW Caravelle	Ford Fiesta	Fiat Uno
205	205 Rallye	Seat Ibiza SXI	AX Sport

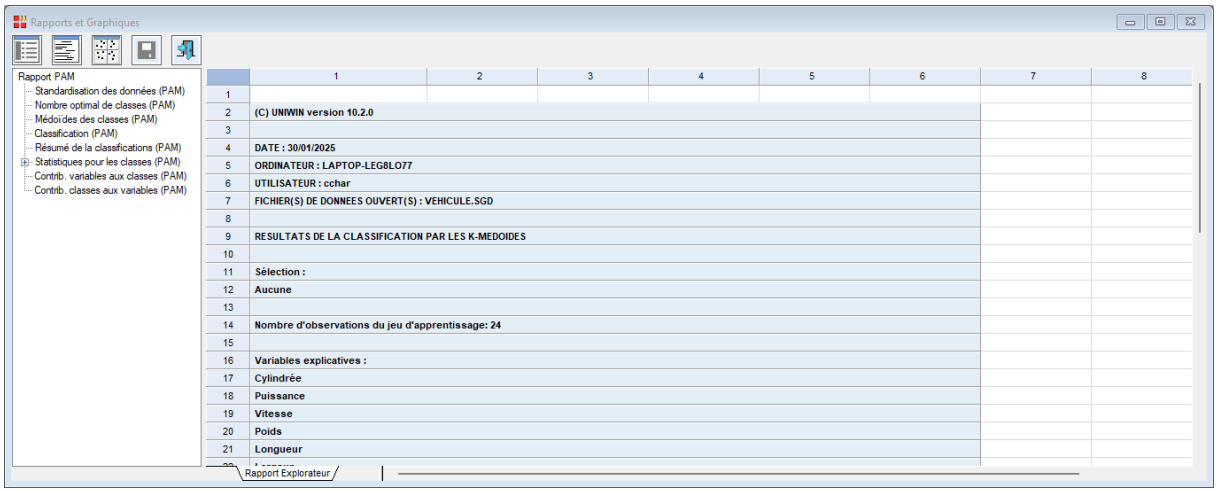
Cliquons sur l'icône KM dans le ruban Décrire et choisissons K-médoïdes. La boîte de dialogue montrée ci-après s'affiche.

Nous choisissons toutes les variables de *Cylindrée* à *Largeur* comme variables quantitatives, la variable *Mesures* comme variable contenant les libellés associés et la variable *Modèle* comme variable contenant les libellés des observations.

Nous demandons de former 3 classes et choisissons de faire une classification sur les données standardisées.



Cliquons sur Ok. UNIWIN débute le calcul de la classification. Après quelques instants, l'écran suivant s'affiche :



Le premier tableau affiche les paramètres de la standardisation des données.

STANDARDISATION DES DONNEES (PAM)

Moyenne et MAD (écart absolu moyen)

	Moyenne	MAD
Cylindrée	1906,12500	405,95833
Puissance	113,66667	31,11111
Vitesse	183,08333	18,67361
Poids	1110,83333	197,01389
Longueur	421,58333	35,60417
Largeur	168,83333	5,87500

Le deuxième tableau affiche pour les nombres de classes variant de 2 au nombre maximum de classes indiqué (10), les coefficients moyens de silhouette. Ces coefficients sont également représentés de façon graphique. Ils aident à déterminer le nombre adéquat de classes à former.

DETERMINATION DU NOMBRE OPTIMAL DE CLASSES (PAM)

Silhouette = coefficient moyen de silhouette

	Silhouette
2 classes	0,40213
3 classes	0,38557
4 classes	0,34774
5 classes	0,30458
6 classes	0,34514
7 classes	0,32795
8 classes	0,35126
9 classes	0,31775
10 classes	0,27889

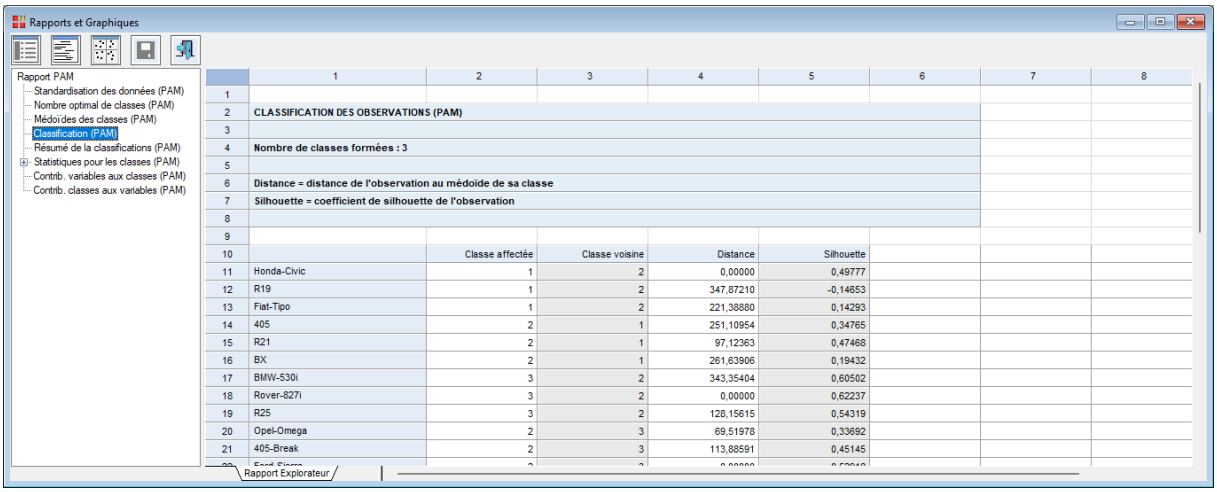
Le troisième tableau affiche les médoïdes des classes formées.

MEDOIDES DES CLASSES (PAM)

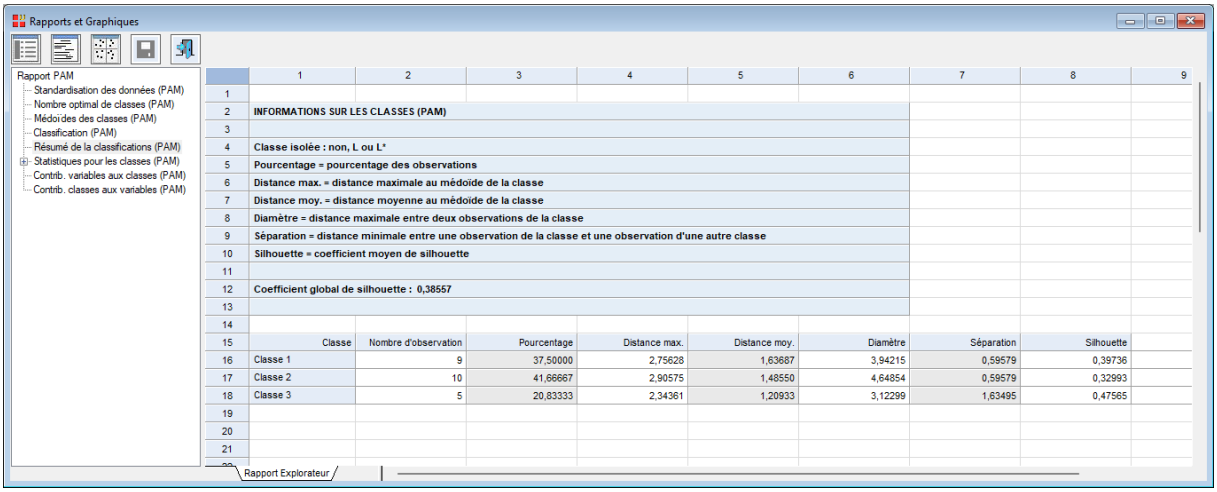
	Observation	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur
Classe 1 - médoïde = Honda-Civic	1	1396	90	174	850	369	166
Classe 2 - médoïde = Ford-Sierra	12	1993	115	185	1190	451	172
Classe 3 - médoïde = Rover-827i	8	2675	177	222	1365	469	175

La classe 1 est représentée par l'observation 1, la classe 2 par l'observation 12 et la classe 3 par l'observation 8.

Le quatrième tableau affiche pour chaque observation sa classe d'affectation, sa classe voisine, sa distance au médoïde et son coefficient de silhouette.



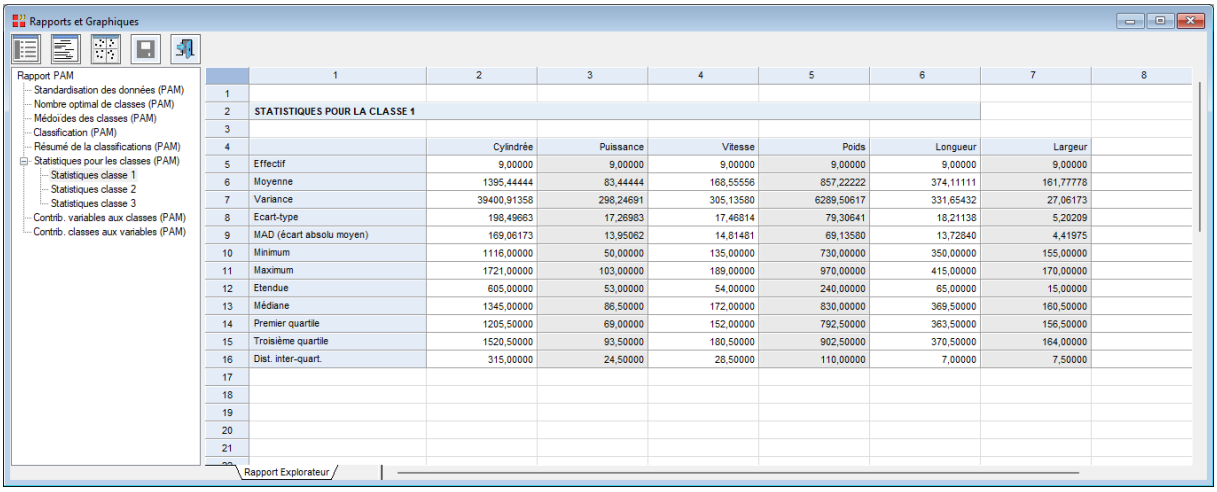
Le cinquième tableau affiche un résumé de la classification.



- Classe isolée (non isolée, L ou L*)
- Pourcentage des observations
- Distance maximale au médoïde de la classe
- Distance moyenne au médoïde de la classe
- Diamètre de la classe : distance maximale entre deux observations de la classe
- Séparation de la classe : distance minimale entre une observation de la classe et une observation d'une autre classe
- Coefficient moyen de silhouette

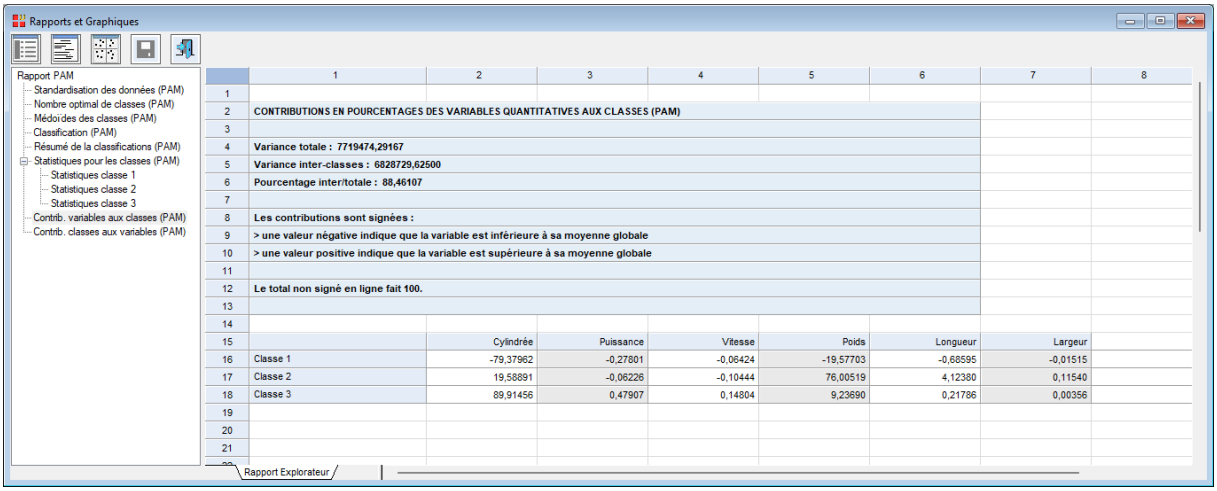
Une classe est une classe L* si son diamètre est plus petit que sa séparation. Une classe est une classe L si pour chaque observation i la dissimilarité maximale entre i et toute autre observation du cluster est plus petite que la dissimilarité minimale entre i et toute observation d'un autre cluster. Une classe L* est donc également une classe L.

Le tableau suivant affiche pour chaque classe un ensemble de statistiques descriptives pour chacune des variables utilisées pour la classification.

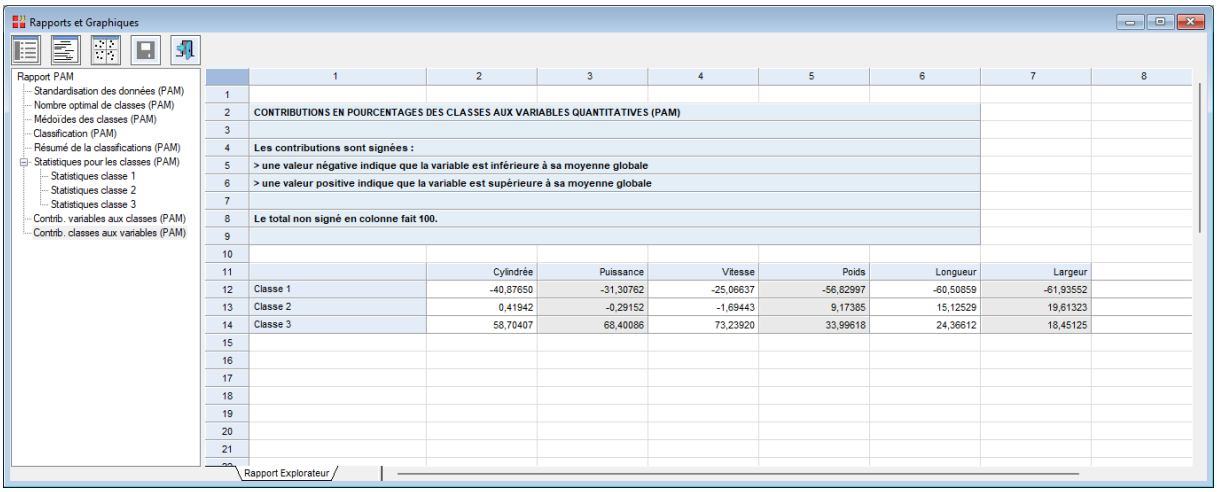


	1	2	3	4	5	6	7	8
1								
2	**STATISTIQUES POUR LA CLASSE 1**							
3								
4		Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur	
5	Effectif	9,00000	9,00000	9,00000	9,00000	9,00000	9,00000	
6	Moyenne	1395,44444	83,44444	168,55556	857,22222	374,11111	161,77778	
7	Variance	39400,91358	298,24691	305,13580	6289,50617	331,65432	27,06173	
8	Ecart-type	198,49663	17,26983	17,46814	79,30641	18,21138	5,20209	
9	MAD (écart absolu moyen)	169,06173	13,95062	14,81481	69,13580	13,72840	4,41975	
10	Minimum	1116,00000	50,00000	135,00000	730,00000	350,00000	155,00000	
11	Maximum	1721,00000	103,00000	189,00000	970,00000	415,00000	170,00000	
12	Etendue	605,00000	53,00000	54,00000	240,00000	65,00000	15,00000	
13	Médiane	1345,00000	86,50000	172,00000	830,00000	369,50000	160,50000	
14	Premier quartile	1205,50000	69,00000	152,00000	792,50000	363,50000	156,50000	
15	Troisième quartile	1520,50000	93,50000	180,50000	902,50000	370,50000	164,00000	
16	Dist. inter-quart.	315,00000	24,50000	28,50000	110,00000	7,00000	7,50000	
17								
18								
19								
20								
21								

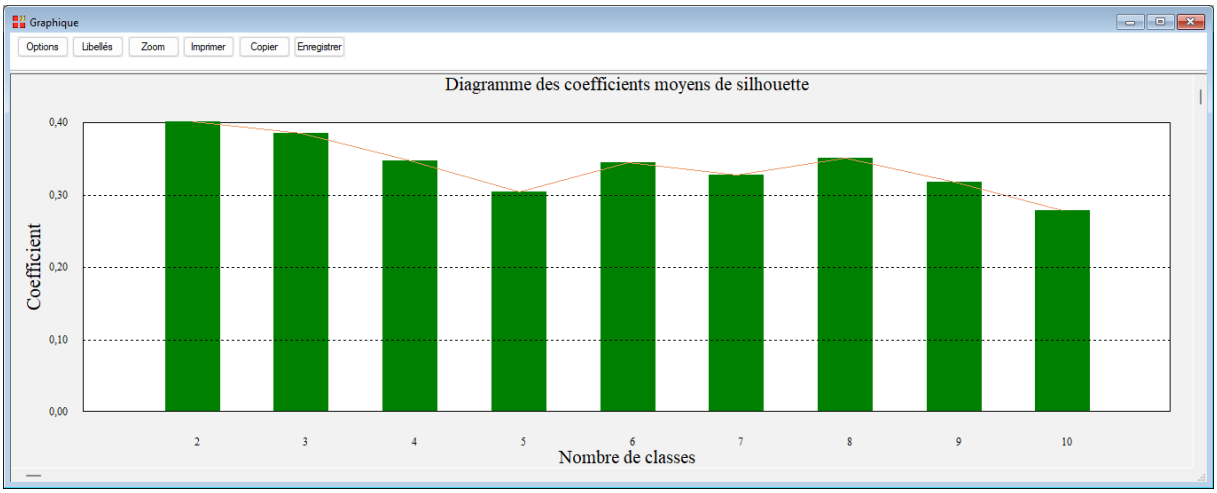
Les deux tableaux suivants affichent les contributions signées en pourcentages des variables aux classes et les contributions signées en pourcentages des classes aux variables.



	1	2	3	4	5	6	7	8
1								
2	**CONTRIBUTIONS EN POURCENTAGES DES VARIABLES QUANTITATIVES AUX CLASSES (PAM)**							
3								
4	**Variance totale : 7719474,29167**							
5	**Variance inter-classes : 6828729,62500**							
6	**Pourcentage inter/totale : 88,46107**							
7								
8	**Les contributions sont signées :**							
9	**> une valeur négative indique que la variable est inférieure à sa moyenne globale**							
10	**> une valeur positive indique que la variable est supérieure à sa moyenne globale**							
11								
12	**Le total non signé en ligne fait 100.**							
13								
14								
15		Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur	
16	Classe 1	-79,37962	-0,27801	-0,06424	-19,57703	-0,68595	-0,01515	
17	Classe 2	19,58891	-0,06226	-0,10444	76,00519	4,12380	0,11540	
18	Classe 3	89,91456	0,47907	0,14804	9,23690	0,21786	0,00356	
19								
20								
21								

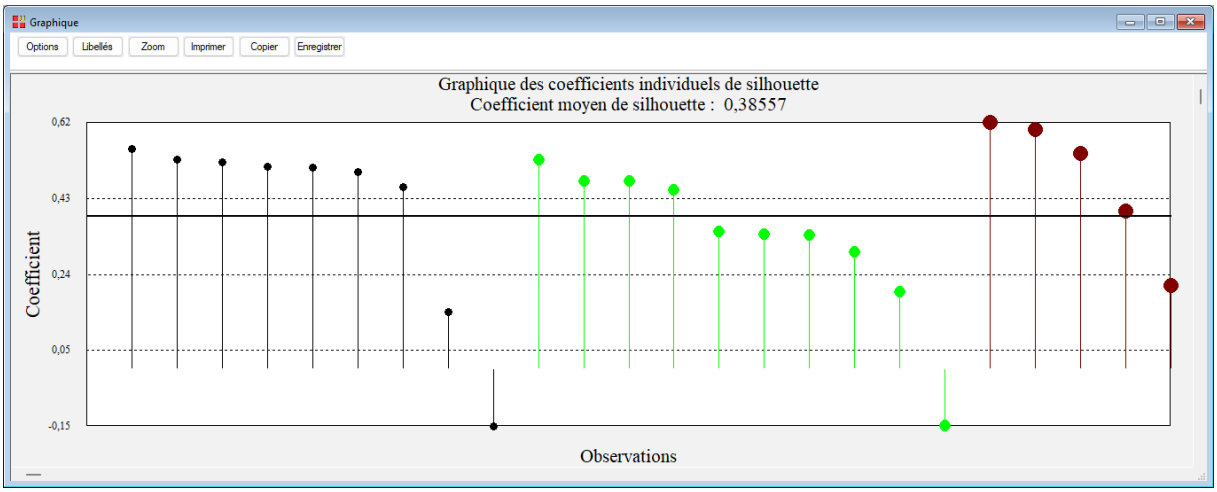


Le diagramme des coefficients moyens de silhouette permet de visualiser l'évolution de ces coefficients en fonction du nombre de classes formées.



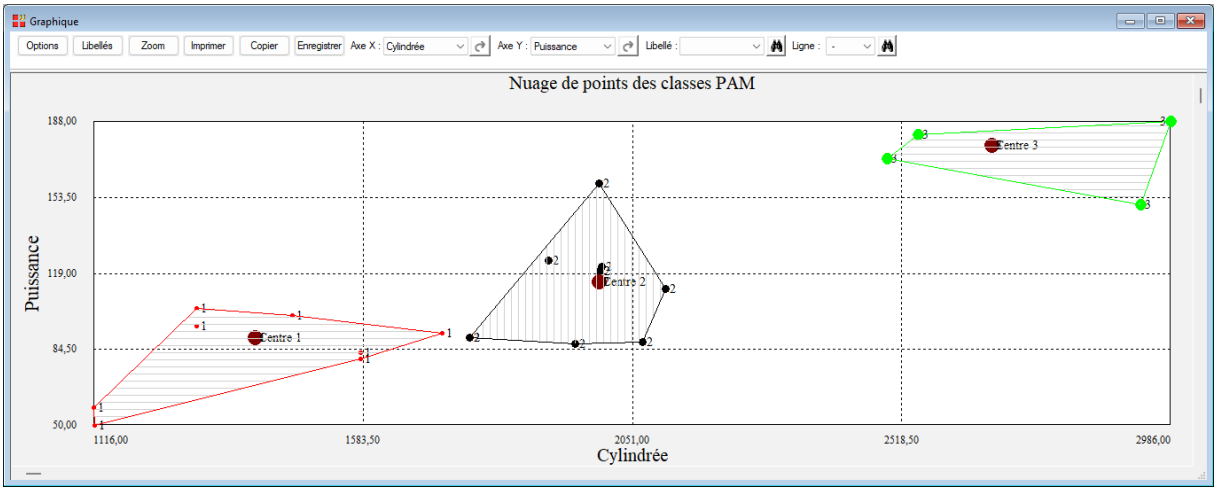
Le graphique des coefficients individuels de silhouette permet de visualiser ces coefficients pour chacune des observations. Pour chaque observation, le coefficient de silhouette est la différence entre la distance moyenne avec les observations de sa classe et la distance moyenne avec les observations des autres classes. Il varie entre -1 et +1. Si négatif, l'observation est en moyenne plus proche de la classe voisine que de la sienne et donc elle est donc mal classée. Si positif, l'observation est en moyenne plus proche de sa classe que de la classe voisine et donc elle est donc bien classée.

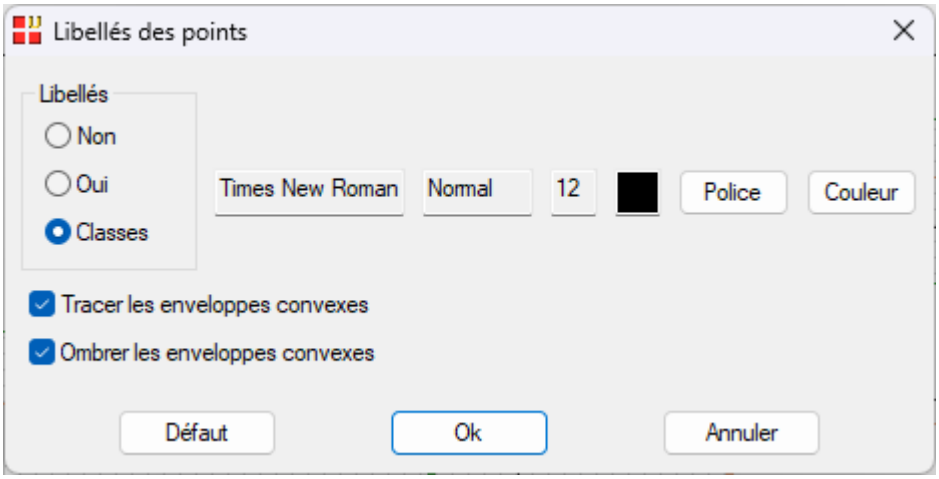
La ligne horizontale indique la position du coefficient moyen de silhouette. Si la plupart des observations d'une classe ont des coefficients inférieurs à ce coefficient moyen, cela indique probablement que la partition obtenue n'est pas bonne.



Le graphique du nuage de points des classes permet de visualiser les classes formées par rapport à deux variables quantitatives sélectionnées.

Le bouton ‘Libellés’ permet de préciser les libellés affichés et si les enveloppes convexes des classes sont tracées ou non.



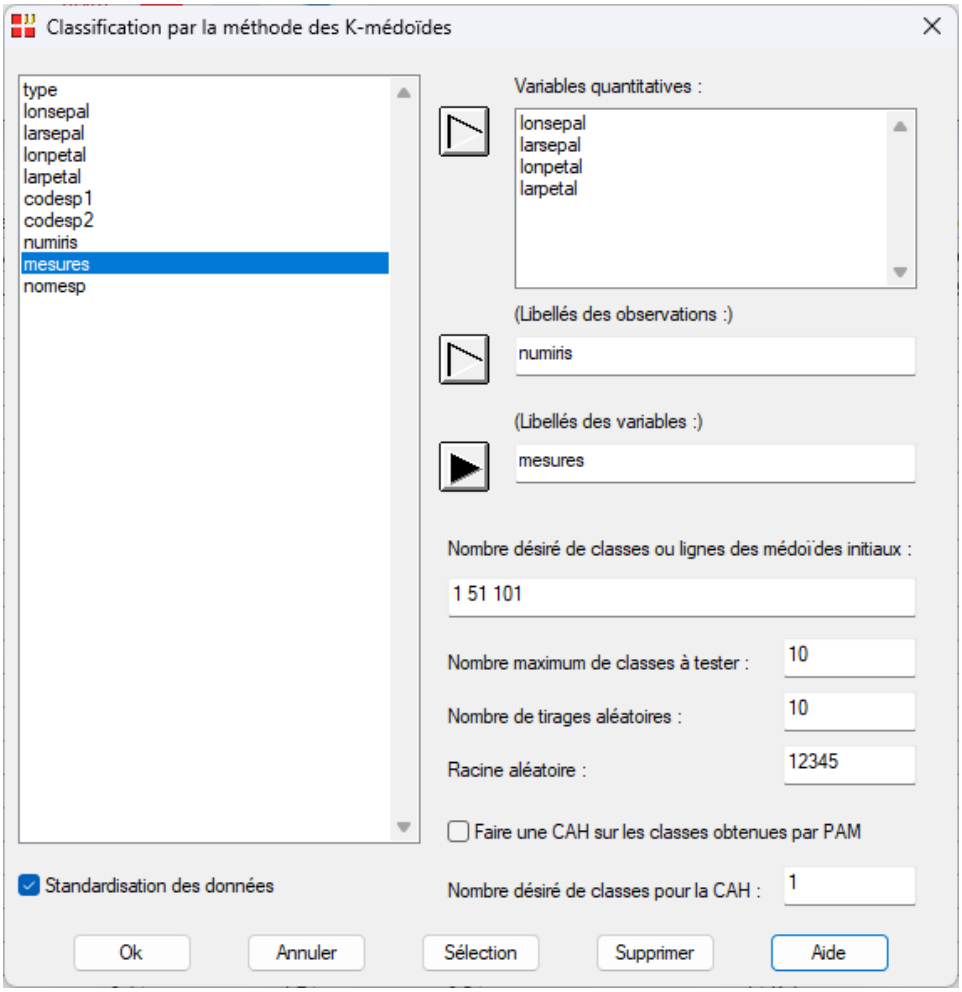


Exemple 3 : Fichier IRIS3

Nous utiliserons le fichier IRIS3 pour illustrer ce troisième exemple.

Ce fichier contient pour 150 iris les mesures des quatre caractéristiques suivantes exprimées en millimètres : longueur du sépale, largeur du sépale, longueur du pétale et largeur du pétale

Renseignons la boîte de dialogue comme montré ci-dessous en précisant les médoïdes initiaux à utiliser : les lignes 1, 51 et 101 du fichier des données.

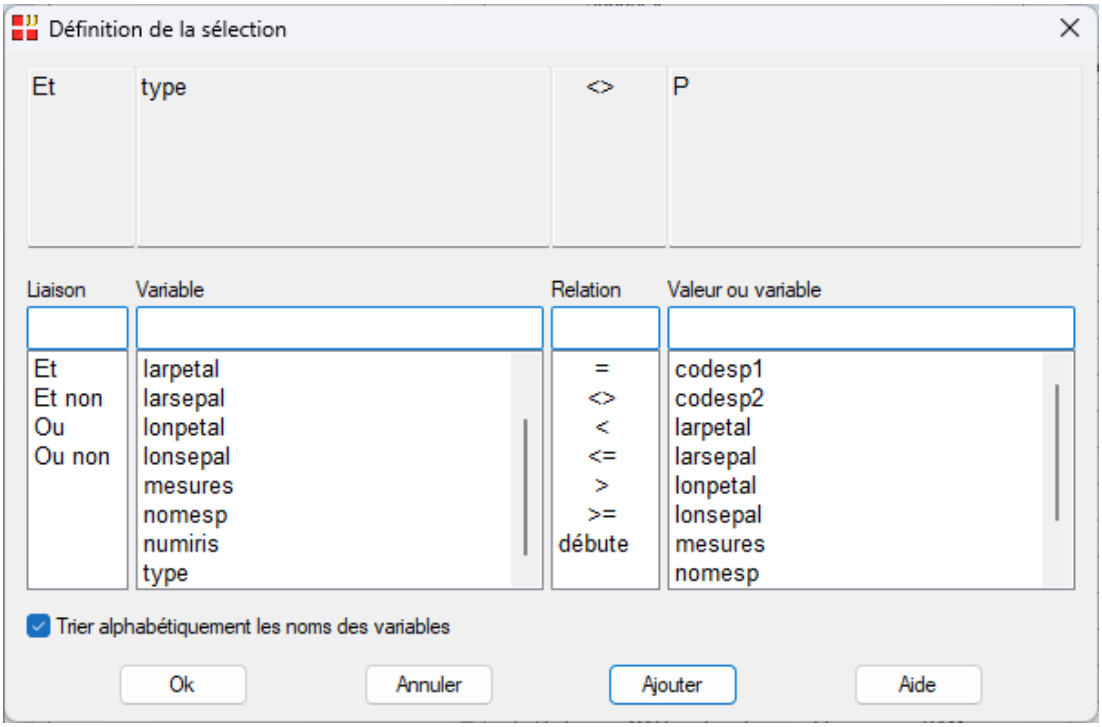


Cliquons sur le bouton ‘Sélection’ pour sélectionner les observations à utiliser comme jeu d’apprentissage.

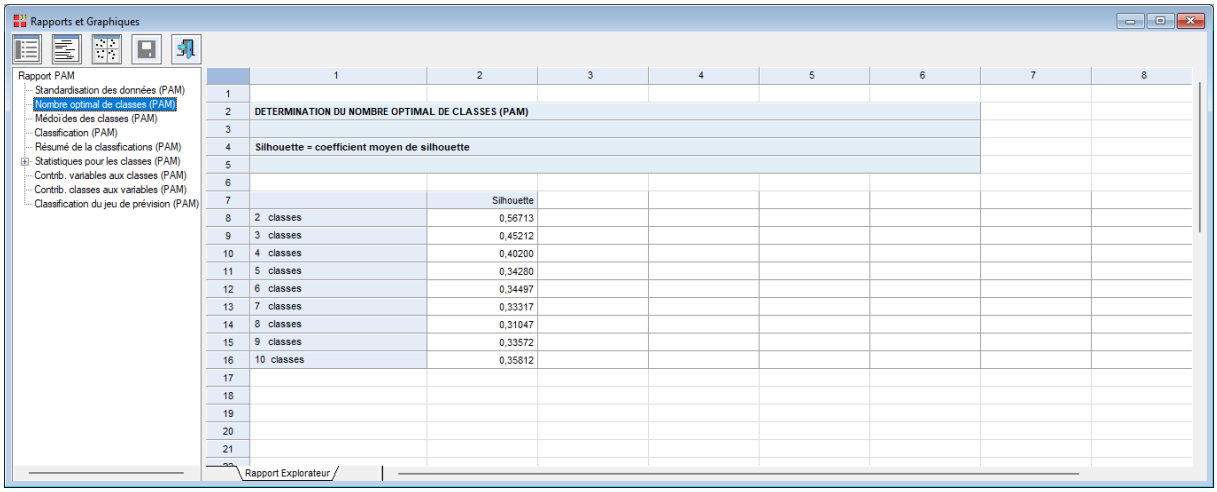
Un message nous indique que 144 observations seront utilisées.

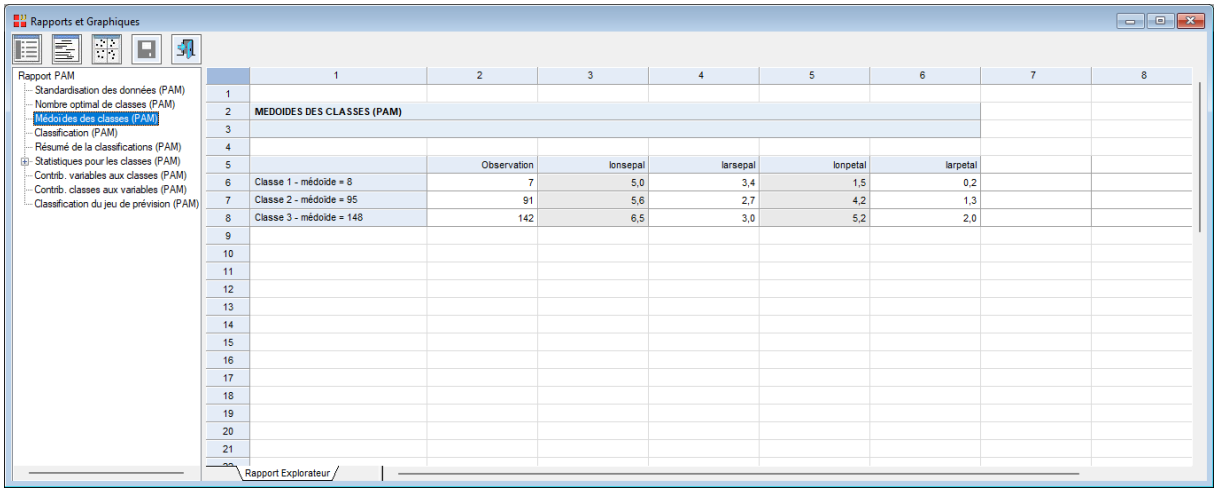
Cliquons sur Ok.

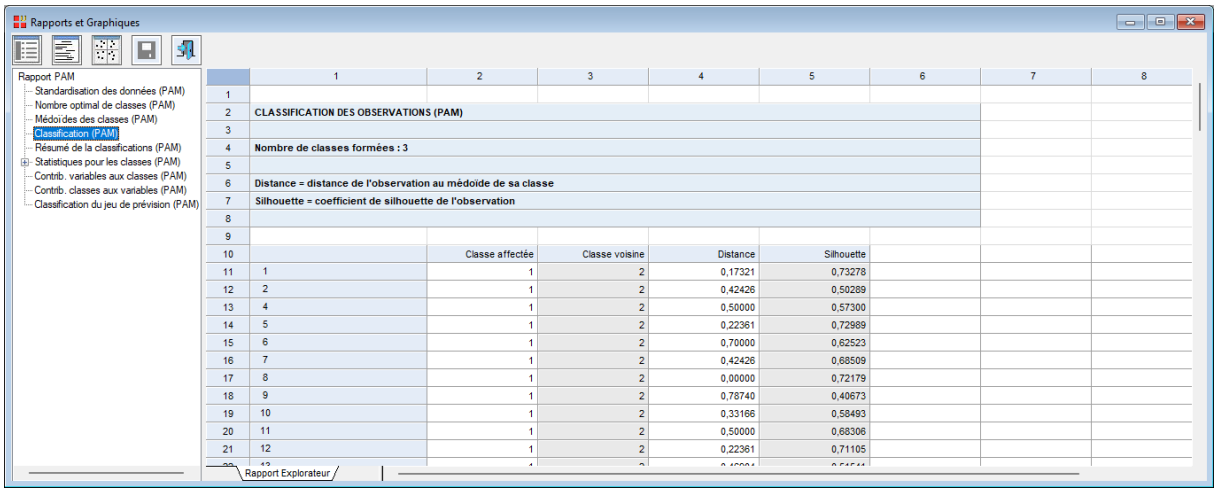
Un message nous indique que les lignes non sélectionnées seront utilisées comme jeu de prévision.

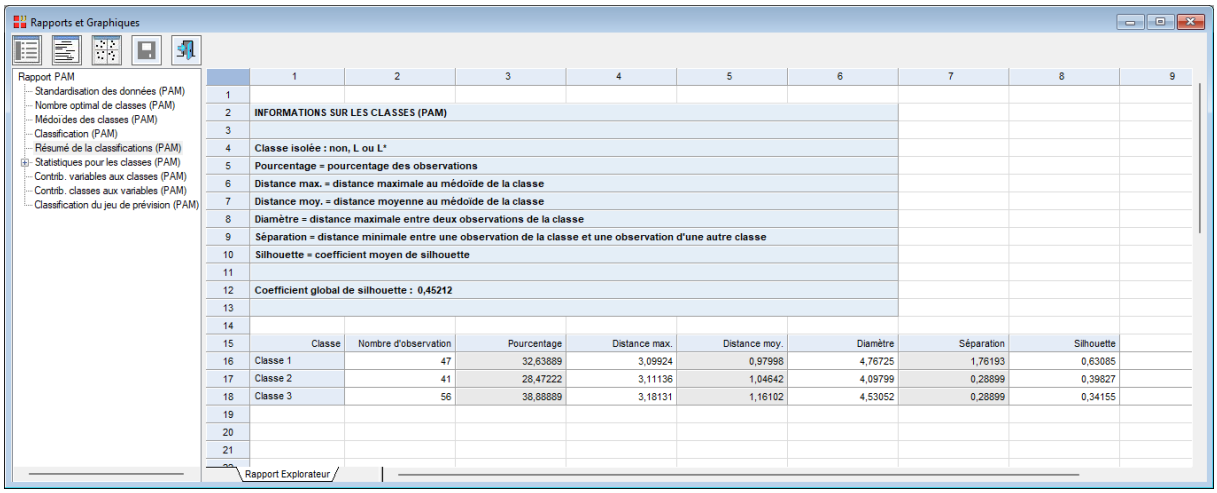


Visualisons quelques résultats obtenus.

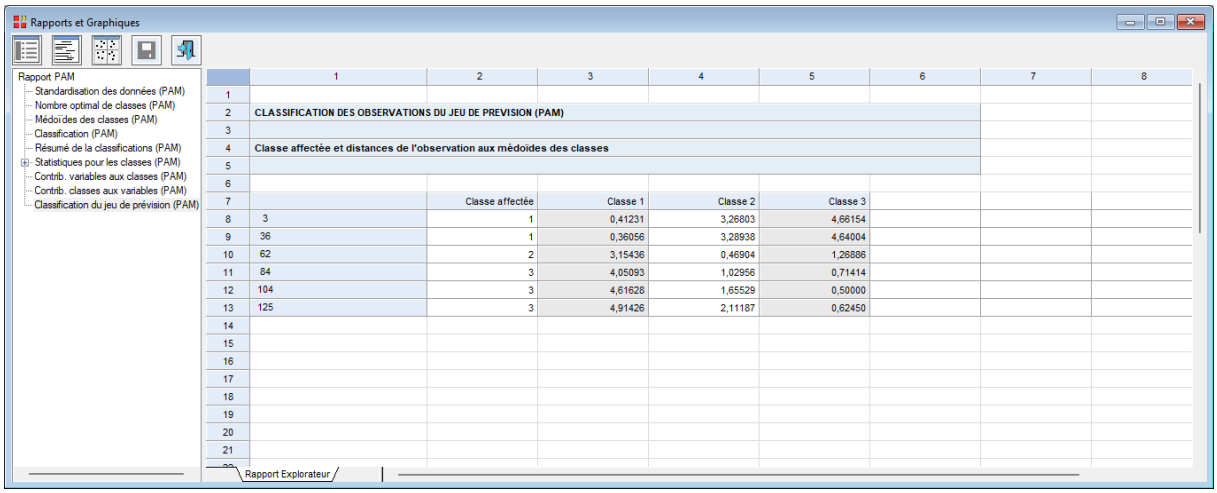


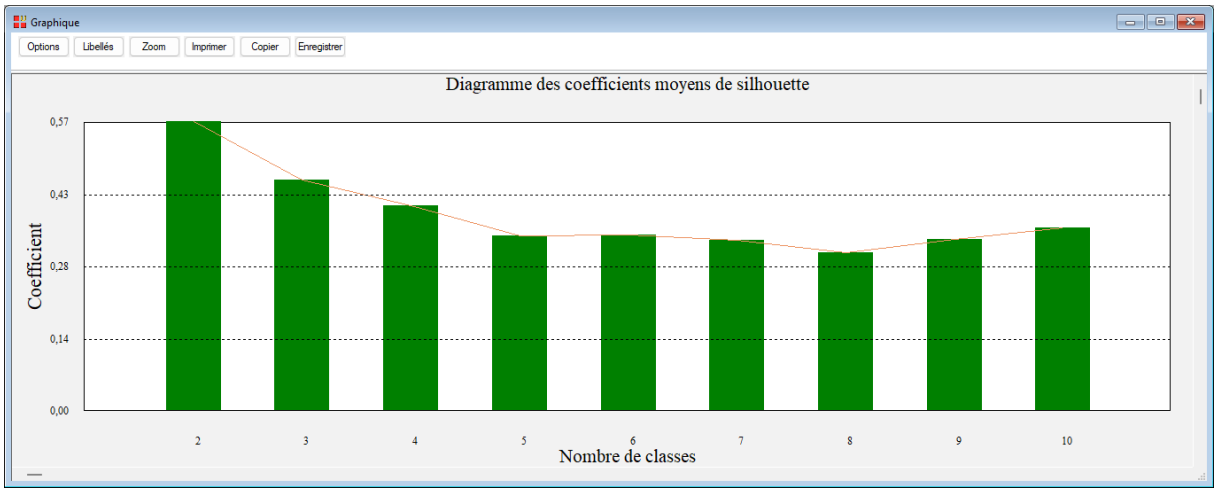


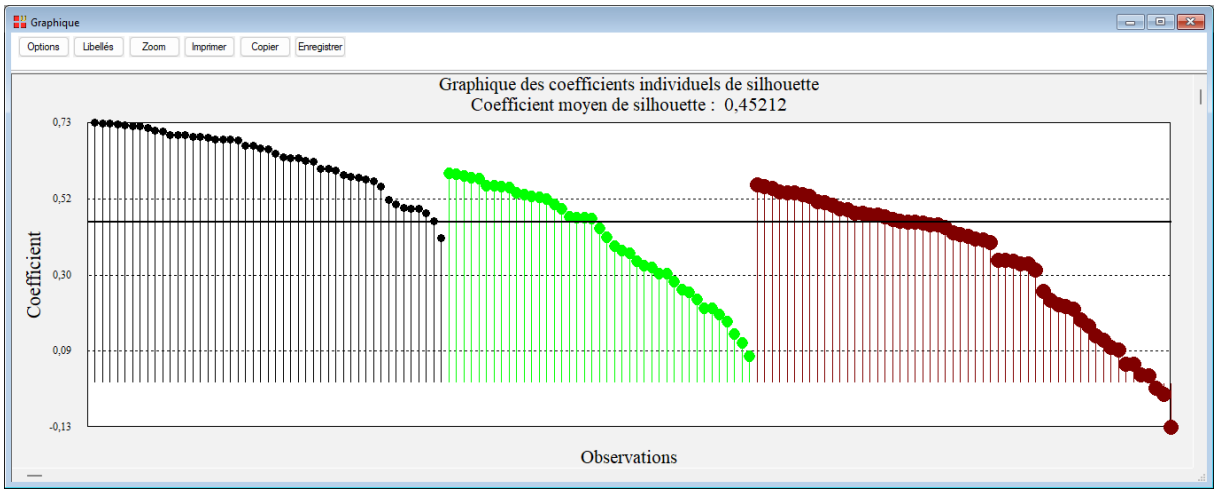


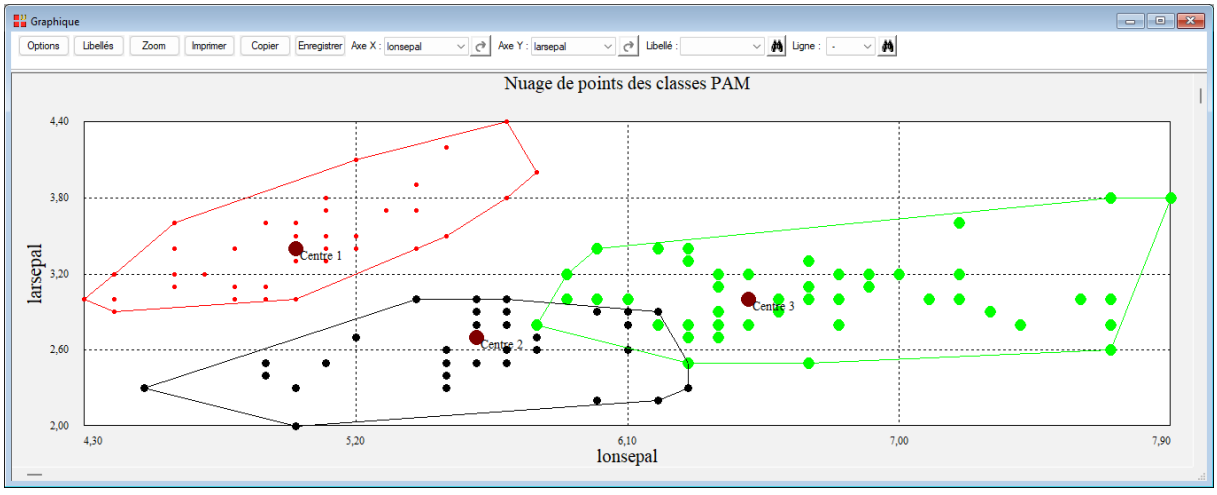


Un tableau indique les classes affectées aux données du jeu de prévision :




Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Diagramme des coefficients moyens de silhouette
Coefficient
0,57
0,43
0,28
0,14
0,00
2
3
4
5
6
7
8
9
10
Nombre de classes


Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Graphique des coefficients individuels de silhouette
Coefficient moyen de silhouette : 0,45212
Coefficient
0,73
0,52
0,30
0,09
-0,13
Observations

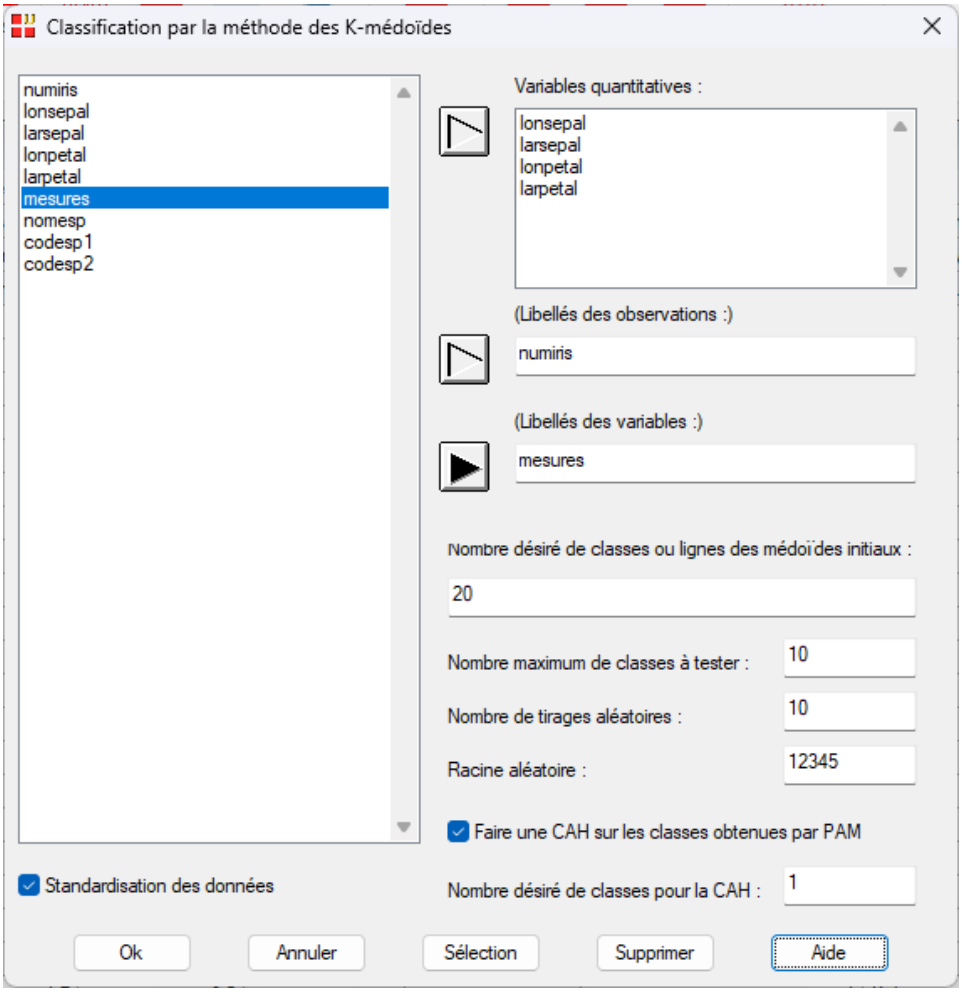

Graphique
Options
Libellés
Zoom
Imprimer
Copier
Enregistrer
Axe X :
lonsepal
Axe Y :
larsepal
Libellé :
Ligne :
Nuage de points des classes PAM
larsepal
4,40
3,80
3,20
2,60
2,00
Centre 1
Centre 2
Centre 3
4,30
5,20
6,10
7,00
7,90
lonsepal

Exemple 4 : Fichier IRIS - Classification mixte

Si le nombre d'observations à classer est important, la classification mixte est une démarche intéressante. Elle se déroule en trois étapes :

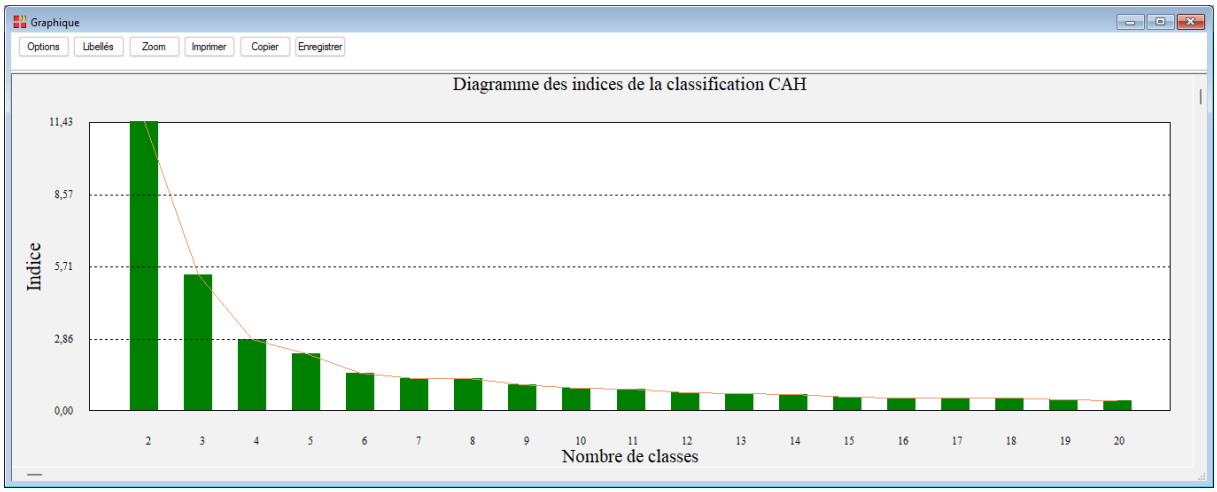
1. Partitionnement en *q* classes par la méthode K-médoïdes
2. Classification ascendante hiérarchique (CAH) sur les *q* classes issues des K-médoïdes avec pondération par les effectifs des classes
3. Partition finale obtenue par troncature de l'arbre CAH

A titre d'exemple illustratif, mettons cette démarche en œuvre en utilisant le fichier IRIS et en demandant un nombre de classes égal à 20 (le nombre maximum de classes à tester est alors automatiquement défini à 20).



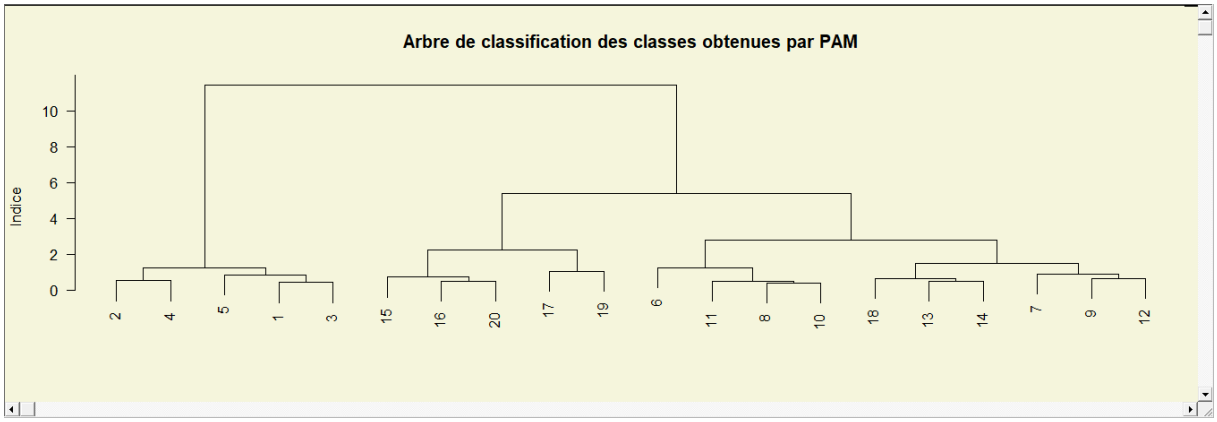
Cochons la case 'Faire une classification CAH sur les classes obtenues par PAM et demandons 1 classe pour la CAH.

Affichons le diagramme des indices de la classification CAH.

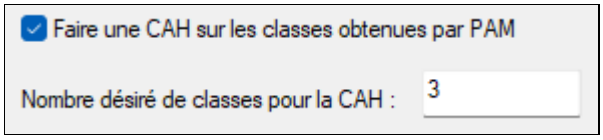


Ce graphique nous indique que 3 classes est un choix possible.

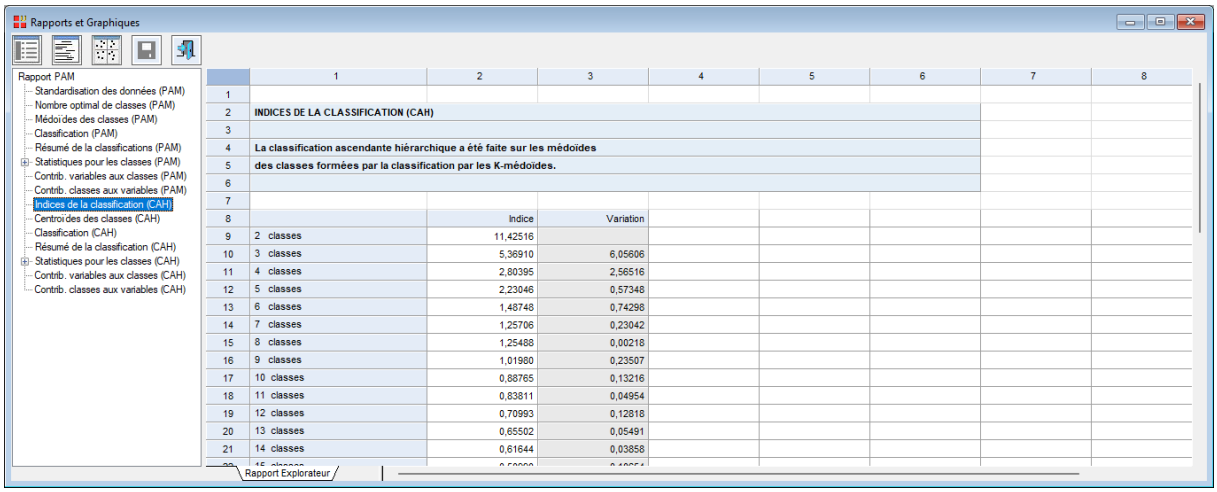
Visualisons l'arbre de la classification.

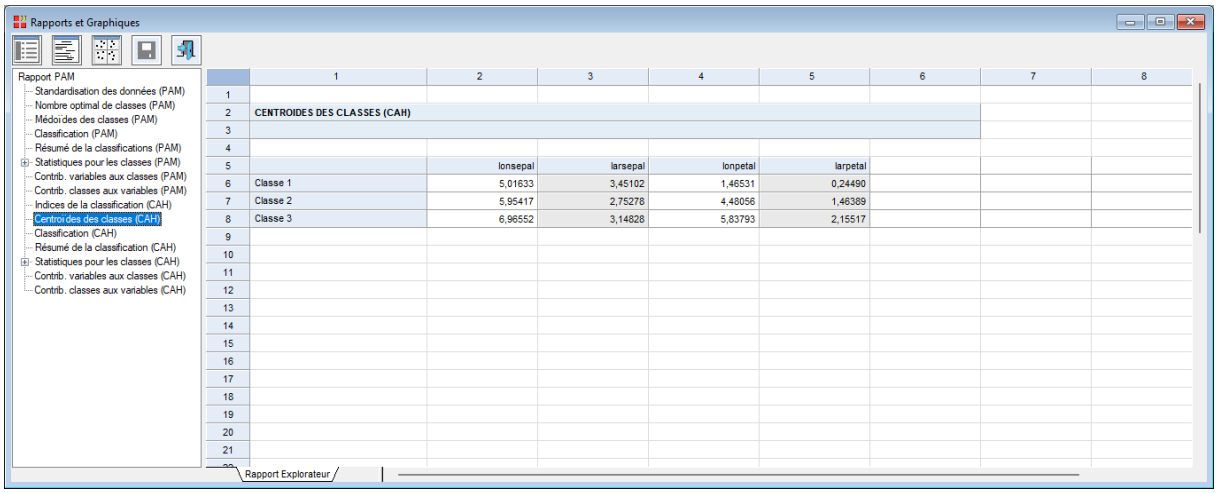


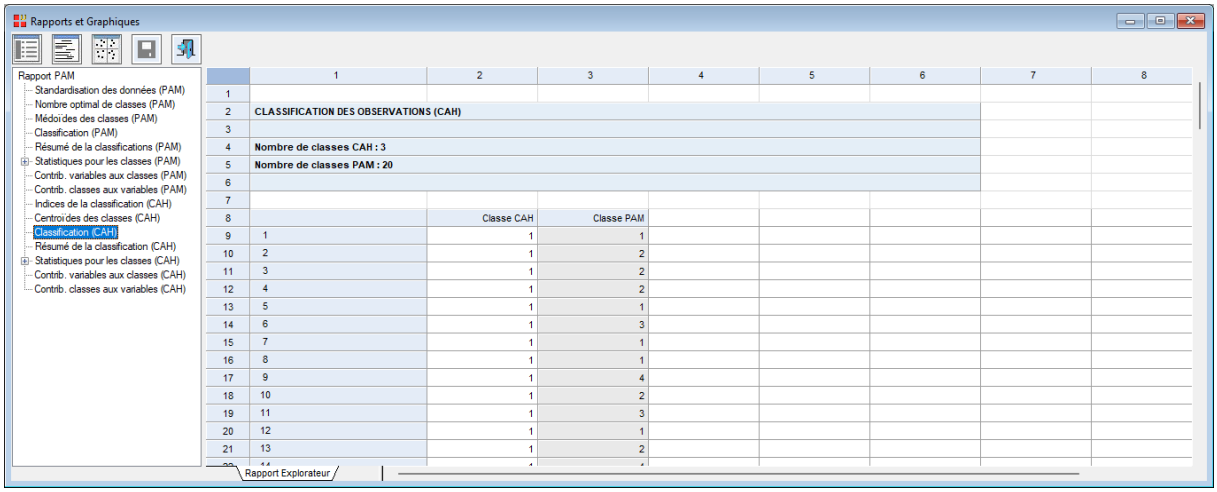
Revenons à la boîte de dialogue d'entrée des données en cliquant sur  et indiquons un nombre de classes égal à 3 pour la classification CAH.

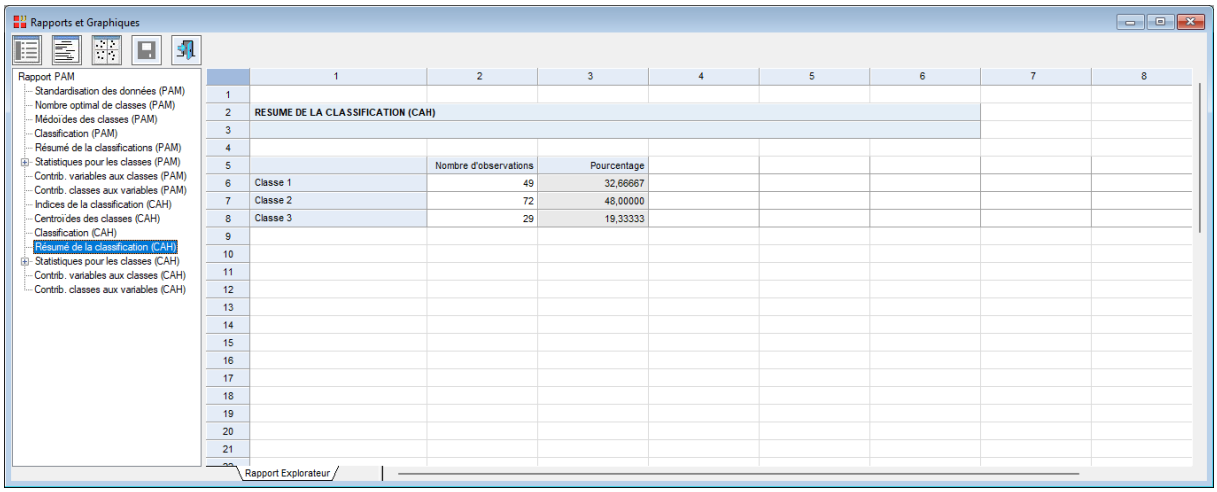


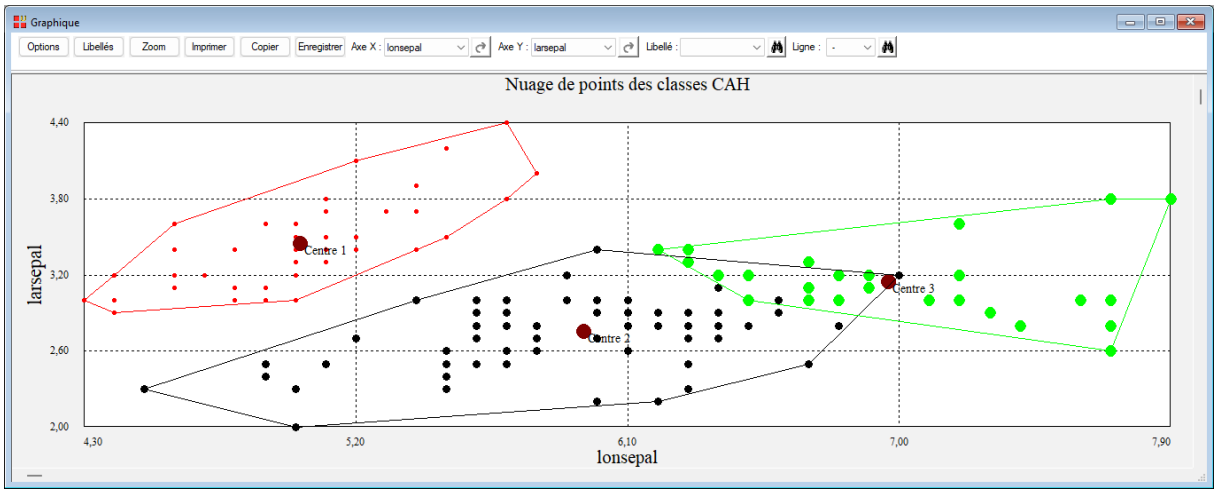
Exécutons à nouveau l'analyse et visualisons les résultats obtenus dans le rapport ainsi que le graphique du nuage des points des classes de la CAH.


Rapports et Graphiques
Rapport PAM
Standardisation des données (PAM)
Nombre optimal de classes (PAM)
Médoïdes des classes (PAM)
Classification (PAM)
Résumé de la classifications (PAM)
Statistiques pour les classes (PAM)
Contrib. variables aux classes (PAM)
Contrib. classes aux variables (PAM)
Indices de la classification (CAH)
Centroïdes des classes (CAH)
Classification (CAH)
Résumé de la classification (CAH)
Statistiques pour les classes (CAH)
Contrib. variables aux classes (CAH)
Contrib. classes aux variables (CAH)
INDICES DE LA CLASSIFICATION (CAH)
La classification ascendante hiérarchique a été faite sur les médoïdes
des classes formées par la classification par les K-médoïdes.
Indice
Variation
2 classes 11,42516
3 classes 5,36910 6,05606
4 classes 2,80395 2,56516
5 classes 2,23046 0,57348
6 classes 1,48748 0,74298
7 classes 1,25706 0,23042
8 classes 1,25488 0,00218
9 classes 1,01980 0,23507
10 classes 0,88765 0,13216
11 classes 0,83811 0,04954
12 classes 0,70993 0,12818
13 classes 0,65502 0,05491
14 classes 0,61644 0,03858
Rapport Explorateur

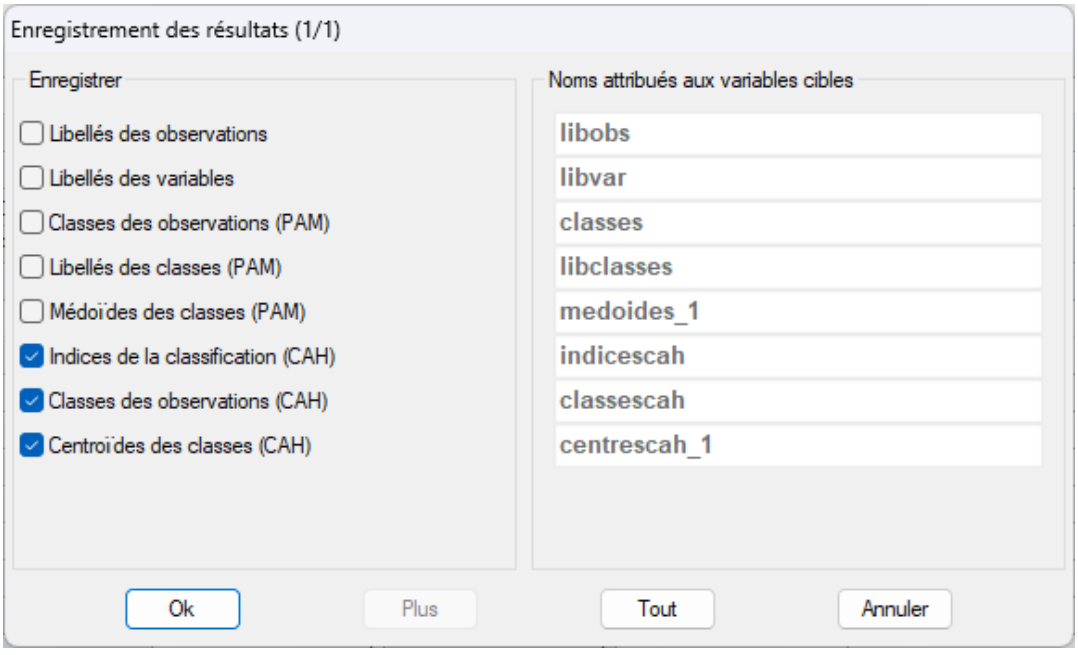

Rapports et Graphiques
Rapport PAM
Standardisation des données (PAM)
Nombre optimal de classes (PAM)
Médoïdes des classes (PAM)
Classification (PAM)
Résumé de la classifications (PAM)
Statistiques pour les classes (PAM)
Contrib. variables aux classes (PAM)
Contrib. classes aux variables (PAM)
Indices de la classification (CAH)
Centroïdes des classes (CAH)
Classification (CAH)
Résumé de la classification (CAH)
Statistiques pour les classes (CAH)
Contrib. variables aux classes (CAH)
Contrib. classes aux variables (CAH)
CENTROIDES DES CLASSES (CAH)
lonsepal larsepal lonpetal larpetal
Classe 1 5,01633 3,45102 1,46531 0,24490
Classe 2 5,95417 2,75278 4,48056 1,46389
Classe 3 6,96552 3,14828 5,83793 2,15517
Rapport Explorateur


Rapports et Graphiques
Rapport PAM
Standardisation des données (PAM)
Nombre optimal de classes (PAM)
Médoïdes des classes (PAM)
Classification (PAM)
Résumé de la classifications (PAM)
Statistiques pour les classes (PAM)
Contrib. variables aux classes (PAM)
Contrib. classes aux variables (PAM)
Indices de la classification (CAH)
Centroïdes des classes (CAH)
Classification (CAH)
Résumé de la classification (CAH)
Statistiques pour les classes (CAH)
Contrib. variables aux classes (CAH)
Contrib. classes aux variables (CAH)
CLASSIFICATION DES OBSERVATIONS (CAH)
Nombre de classes CAH : 3
Nombre de classes PAM : 20
Classe CAH Classe PAM
1 1 1
2 1 2
3 1 2
4 1 2
5 1 1
6 1 3
7 1 1
8 1 1
9 1 4
10 1 2
11 1 3
12 1 1
13 1 2
Rapport Explorateur





Il est possible d'enregistrer ces résultats :



Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure.

Variable	*Contenu*
libobs	Libellés des observations du jeu d'apprentissage
libvar	Libellés des variables
classes	Classes des observations du jeu d'apprentissage (PAM)
libclasses	Libellés des classes (PAM)
medoides	Médoïdes des classes (PAM)
libobsprev	Libellés des observations du jeu de prévision (PAM)
classesprev	Classes affectées aux observations du jeu de prévision (PAM)
indicescah	Indices de la classification CAH (classification mixte)
classescah	Classes des observations CAH (classification mixte)
centrescah	Centroïdes non standardisés des classes CAH (classification mixte)

Références

Documentation du package R 'stats' (2024)

https://rdrr.io/r/stats/stats-package.html

Documentation du package R 'cluster' (2024)

https://cran.r-project.org/web/packages/cluster/cluster.pdf